DOCUMENT RESUME

ED 404 179 SE 059 721

AUTHOR Sugrue, Brenda; And Others
TITLE Mapping Test Items to the 1992 NAEP Mathematics Achievement Level Descriptions: Mathematics Educators' Interpretations and Their Relationship to Student Performance.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO CSE-TR-393
PUB DATE Feb 95
CONTRACT RS90159001
NOTE 147p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; *Knowledge Representation; Mathematics Achievement; *Mathematics Tests; National Competency Tests; Standards; *Student Evaluation; Test Construction; *Test Items; *Test Validity
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

This study aims to evaluate the degree to which the achievement level descriptions adopted by the National Assessment Governing Board (NAGB) for the 1992 assessment in mathematics accurately represent what students at a given achievement level can do. NAGB descriptions of the levels were used to form lists of statements about what students at a given grade and level should be able to do. Judges then used those statements to identify items from the 1992 mathematics assessment that called for the knowledge, skill, or understanding contained in the descriptor-based statements. The three main findings that emerge from this study are: (1) the achievement level descriptions are not clear enough to support consistent interpretation; (2) the 1992 National Assessment of Educational Progress (NAEP) mathematics assessment provided sparse coverage or no coverage of some of the skills included in the achievement level descriptions; and (3) frequently many, in some cases a majority, of the students at a given level did not successfully answer items linked to certain aspects of the descriptions at that level. In summary, the analyses do not support the validity of the published narrative descriptions as characterizations of what students within specified scoring ranges can do. Contains 19 references. (DDR)

C·R·E·S·S·T

National Center for Research on Evaluation, Standards, and Student Testing

Mapping Test Items to the 1992 NAEP Mathematics Achievement Level Descriptions: Mathematics Educators' Interpretations and Their Relationship to Student Performance

CSE Technical Report 393

Brenda Sugrue, John Novak, Leigh Burstein,
CRESST/University of California, Los Angeles

Elizabeth Lewis, Daniel M. Koretz
CRESST/ RAND

Robert L. Linn
CRESST/University of Colorado at Boulder

- **UCLA Center for the Study of Evaluation**

in collaboration with:

- University of Colorado
- NORC, University of Chicago
- LRDC, University of Pittsburgh
- University of California, Santa Barbara
- University of Southern California
- The RAND Corporation

Mapping Test Items to the 1992 NAEP Mathematics Achievement Level Descriptions: Mathematics Educators' Interpretations and Their Relationship to Student Performance

CSE Technical Report 393

Brenda Sugrue, John Novak, Leigh Burstein,
CRESST/University of California, Los Angeles

Elizabeth Lewis, Daniel M. Koretz
CRESST/ RAND

Robert L. Linn
CRESST/University of Colorado at Boulder

February 1995

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported in part under the National Center for Education Statistics Contract No. RS90159001 as administered by the U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the National Center for Education Statistics or the U.S. Department of Education.

Contents

Executive Summary

The National Center for Education Statistics (NCES), concerned that the reporting of the National Assessment of Educational Progress (NAEP) should be accurate and informative, asked the NAEP Technical Review Panel (TRP) to evaluate the degree to which the achievement level descriptions adopted by the National Assessment Governing Board (NAGB) for the 1992 assessment in mathematics accurately represent what students at a given achievement level can do. One of the studies conducted as part of that evaluation was designed to answer three questions:

1. **How clear are the achievement level descriptions?**
2. **To what extent is the existing NAEP item pool in mathematics appropriate for generating scores that can be referenced to the knowledge and skills articulated in the achievement level descriptions?**
3. **Are the score ranges for items associated with the achievement levels consistent with the narrative descriptions of their corresponding levels?**

The study was deemed necessary for a number of reasons. First, the clarity and interpretability of the descriptions had not been empirically verified prior to their use. If these descriptions of the capabilities of students are to act as standards against which student performance can be measured, it is important to establish that they convey a clear depiction of the knowledge and skills that distinguish among students scoring at particular levels of the NAEP scale. The description of each level should be clear enough to activate common interpretation, and the differences between the skills that define any two levels should be easily identifiable.

Second, although the process that generated the achievement levels and their descriptions involved a review of the mathematics test item pool, no estimate was made of the extent to which the 1992 item pool covers the content and skills included in the final description of each level (Bourque, 1993). If the item pool is not sufficient to provide indicators of students' knowledge and

skills in the areas mentioned in the operational definitions of the levels, then it would not be appropriate to interpret students' scores on the 1992 NAEP in mathematics in terms of narrative descriptions.

Third, the process of setting the 1992 score ranges for the levels was not validated by reviewing the performance of students in those score ranges on items that targeted skills included in the descriptions at specific levels. If students who score at particular levels do not perform well on items that are judged to measure knowledge and skills included in the narrative descriptions of those levels, then, again, it would not be appropriate to conclude that students scoring at particular levels know and are able to do what the operational descriptions indicate they should be able to do.

In the study reported here, the NAGB descriptions of the levels were used to form lists of statements about what students at a given level and grade should be able to do. Judges (mathematics educators familiar with the curriculum at the target grade levels) then used those statements (without being told the level from which the statement was taken) to identify items from the 1992 mathematics assessment that called for the knowledge, skill, or understanding contained in the descriptor-based statements. Performance of students on the identified items was then summarized for each level in each grade.

Three main findings emerged from this study:

The results of the mathematics educators' mapping of items to descriptors indicated that

1. **the achievement level descriptions are not clear enough to support consistent interpretation even among experienced mathematics educators; and**
2. **the 1992 NAEP mathematics assessment provided sparse coverage or no coverage of some of the skills included in the achievement level descriptions.**

When student performance on aspects of the descriptions, to which a reasonable number of items were consistently mapped, was examined, it emerged that

3. **frequently, many—in some cases, a majority—of the students at a given level did not successfully answer items linked to certain aspects of the descriptions at that level.**

There was considerable variation in judges' application of some descriptors and these descriptors were excluded from the analysis of the distribution of items across descriptors and levels. The descriptors that do not include a reference to specific mathematics content were least consistently mapped to items. The ambiguity of these descriptors would need to be reduced to promote more consistent interpretations among either mathematics educators or the lay public.

In addition, there are many cases where descriptors from different levels have very similar wording. As a consequence of this similarity, many items were mapped to descriptors from more than one level. A clearer distinction between the skills listed in descriptions of different levels would need to be made to enable even mathematics educators to interpret correctly and consistently the intended differences among levels.

The sparse coverage of some elements of the achievement level descriptions may be partly a result of the ambiguity of the descriptions. It may be that, if the descriptions were clearer, then judges would agree more on the items that called for those skills and there would be more items mapped to more descriptors.

Among students whose performance reached a given level, performance on items linked to that level varied and was in many cases lower than many people would consider reasonable. For example, in some instances, the median percentage of students answering correctly was less than 50% on items associated with that level. This variation in performance is greatest for items corresponding to Basic level descriptions. It also emerged that, in many cases, the performance of students who scored in one region of the NAEP scale is as high or higher on a set of items that reflect a higher achievement level than on the set of items that reflect the level at which the students scored.

In sum, then, our analyses do not support the validity of the published narrative descriptions as characterizations of what students within specified score ranges can do. The achievement level descriptions lack clarity. Some elements of the descriptions could not be mapped to the NAEP items, partly

because of the ambiguity of the descriptions; and those that could be mapped to NAEP items did not consistently show performance patterns that would support the validity of the descriptions. In our judgment, descriptions of the achievement levels are not informative unless they accurately portray what students at the various levels can do.

It should be pointed out that the achievement level descriptions mix what have been referred to in the NAEP mathematics frameworks as the "content" and "mathematical ability" dimensions of performance. NAEP has never reported scores across items targeting particular mathematical abilities, either within content areas or independent of content areas. By including references to the ability dimensions in the achievement level descriptions, NAEP is perhaps committing itself to analysis and reporting on aspects of performance that may be difficult, if not impossible, to isolate or tie to particular items or sets of items. Our analysis of the achievement level descriptions was premised on the notion that it is indeed possible to identify both the content and mathematical ability features of the test items.

Characterizations of the levels should align with the framework that was used to generate the item pool and with the actual performance of students on the NAEP. In a criterion-referenced system, both the development of items and the interpretation of performance should be driven by the assessment objectives framework, with the objectives serving as the criteria for judging performance (Nitko, 1984). In the assessment framework, NAEP has already identified a multidimensional set of criteria for analyzing and reporting performance (Educational Testing Service, 1988; Mullis, Dossey, Owen, & Phillips, 1993). A closer alignment of the achievement level descriptions with the assessment framework might considerably improve both the clarity of the descriptions and their reflection in the pool of items generated by the framework. Linking level setting with assessment design from the outset may provide the only means to determine whether it is possible to develop valid descriptions of what students know and can do.

MAPPING TEST ITEMS TO THE 1992 NAEP MATHEMATICS ACHIEVEMENT LEVEL DESCRIPTIONS: MATHEMATICS EDUCATORS' INTERPRETATIONS AND THEIR RELATIONSHIP TO STUDENT PERFORMANCE

Brenda Sugrue, John Novak, Leigh Burstein, Elizabeth Lewis, Daniel M. Koretz, and Robert L. Linn

Background

Since 1984, the National Assessment of Educational Progress (NAEP) has been moving toward more criterion-referenced interpretations of student performance (Phillips et al., 1993). The goal has been to "give meaning to the NAEP scale scores" (Phillips et al., 1993, p. 65) by relating particular score points or regions to descriptions of what students know and can do in the domain of interest. Until recently, reporting of NAEP results was purely descriptive; performance was not compared to standards for what students should be able to do.

The 1988 NAEP reauthorization, however, created the National Assessment Governing Board (NAGB) and gave it the responsibility of identifying appropriate achievement goals for every age and grade level in each subject area. In an effort to meet that responsibility, NAGB has attempted to establish performance standards, called achievement levels (Bourque & Garrison, 1991). Three achievement levels—Basic, Proficient, and Advanced—were established for each grade tested in the 1990 mathematics assessment. However, the validity of these achievement levels was questioned (Linn, Koretz, Baker, & Burstein, 1991; Stufflebeam, Jaeger, & Scriven, 1991). In response to the criticism, a new effort was undertaken by NAGB to establish achievement levels for the 1992 mathematics assessment.

The National Center for Education Statistics (NCES) asked the NAEP Technical Review Panel (TRP) to evaluate the degree to which the achievement

level descriptions adopted by NAGB for the 1992 assessment in mathematics accurately represent what students at a given achievement level can do. At the same time, two other studies were undertaken by the General Accounting Office (GAO) (1993) and the National Academy of Education (NAE) (1993).

The study reported here was one of three studies conducted by the TRP. The findings of all three TRP studies are summarized in a report by Burstein et al. (1993). This report provides a detailed account of the study that was designed to address the following three questions:

1. **How clear are the proposed achievement level descriptions?** In other words, do the descriptions support consistent interpretation of the knowledge and skills associated with the different levels?

2. **To what extent is the existing NAEP item pool in mathematics appropriate for generating scores that can be referenced to the knowledge and skills articulated in the proposed achievement level descriptions?** In other words, is NAEP currently assessing the knowledge and skills that characterize the expectations or standards embodied in NAGB's descriptions of the achievement levels?

3. **Are the score ranges associated with the achievement levels consistent with the narrative descriptions of the levels?** In other words, can we infer that students scoring in particular regions of the NAEP scale can actually do what the achievement level descriptions say they should be able to do?

The study was deemed necessary for a number of reasons. First, the clarity and interpretability of the descriptions had not been empirically verified prior to their use. The format and language of the achievement level descriptions bear remarkable similarity to the descriptions of 1986 proficiency level scales which were considered by Forsyth (1991) to be too ill-defined to permit valid interpretation of performance. If the 1992 achievement level descriptions are to act as standards against which student performance can be measured, it is important to establish that they convey a clear depiction of the knowledge and skills that distinguish among students scoring at particular levels of the NAEP scale. The description of each level should be clear enough to activate common interpretation, and the differences between the skills that define any two levels should be easily identifiable.

Second, although the process that generated the achievement levels and their descriptions involved a review of the mathematics test item pool, no estimate was made of the extent to which the 1992 item pool covers the content and skills included in the final description of each level[1]. The achievement level descriptions were developed with respect to the content/ability framework of objectives that was used to generate the pool of items, not with respect to the items themselves (Bourque, 1993). It was not until after the descriptions of the levels had been generated that exemplary items were selected to illustrate each level, and even the exemplary items were not linked to specific skills that were part of the descriptions (Bourque, 1993).

Third, the process of setting the 1992 score ranges for the levels was not validated by reviewing the performance of students in those ranges on items that targeted skills included in the descriptions of the levels. In its review of NAGB's approach to establishing standards for student performance in mathematics, the GAO (1993) concluded that, in the absence of evidence that particular NAEP scores indicate the presence of the skills that define the levels, the validity of interpretations of scores based on the 1992 achievement levels is questionable.

To generate the kind of validity evidence called for by the GAO, Reckase (1993) suggested a study in which a domain of items would be produced to correspond to a skill included in the description of a particular achievement level; then that domain of items could be administered to students who scored in the range of that particular achievement level. If students whose NAEP scores are in the range associated with the level perform well on the domain of items that represent the skill included in the description of the level, then it would be appropriate to use the description to refer to students scoring in the particular range of NAEP scores. Reckase (1993) also suggests that there might be a way of conducting a similar study using existing items and data. The study reported here is such a study. Sets of items that mathematics educators identified as calling for the knowledge and skills included in each achievement level description were selected from the 1992 NAEP mathematics item pool. Then, performance on those sets of items was summarized for students scoring within the range of each achievement level.

---

[1] The complete text of the achievement level descriptions is contained in Appendix A.

Methodology

Our investigation of the degree to which the descriptions of the NAEP achievement levels provide a valid indication of the actual performance of students at each of the achievement levels included three necessary steps:

a. the achievement level descriptions had to be analyzed and decomposed in order to facilitate their interpretation and mapping to test items;

b. the test items had to be mapped to the content and skills depicted in the descriptions of the levels; and

c. performance on the sets of items mapped to levels had to be examined for students scoring in the range of each achievement level.

This section describes the methodology used to complete each of these steps. The results are reported in a separate section.

Decomposition of the Achievement Level Descriptions

The NAGB achievement level descriptions (which are reproduced in Appendix A) consist of one or two sentences (shaded in gray in the Appendix) about the general mathematics proficiency expected of students at the level, plus one or two additional paragraphs that describe more specific skills that students scoring at that level should be able to perform. It is stated that skills are cumulative across levels.

A number of possible approaches to mapping sets of test items to the achievement level descriptions were considered. For example, the paragraph descriptions could have been left intact and judges could have been asked to sort the items into three groups, each group consisting of items that represented the knowledge and skills described in one of the achievement level descriptions. However, if this strategy had been used, the basis for the classification of any item by any judge would have been unclear. There are so many different skills described even in one sentence of the description of one achievement level that one might assign an item to a particular level for any number of reasons. Two judges using very different criteria might make the same classification decision.

To direct judges' attention to the specific skills associated with each level, and to provide an empirical basis for selecting sets of items to represent each level, the paragraphs were unpacked or parsed to create a list of the distinct

capabilities that characterize each level. The complete parsed versions of the descriptions appear in Appendix B. For example, two distinct Grade 4 Basic level skills were abstracted from the sentence

> Specifically, 4th grade students performing at the basic level should be able to estimate and use basic facts to perform simple computations with whole numbers.

The two distinct skills are:

> ... ability to use basic facts to perform simple computations with whole numbers

and

> ... ability to estimate with whole numbers.

To provide a mechanism for mapping these skill descriptions to test items, any of these abstracted skills that could be assessed with a single test item were compiled into instruments that judges used to identify items that called for particular skills. Each descriptor was prefaced with the phrase "The item calls for." To facilitate efficient item-descriptor matching, the descriptors related to similar content areas were grouped together, regardless of achievement level. Descriptors that related to aspects of the same content, for example, whole numbers or geometry, were subsumed under a higher level descriptor that asked whether the item involved that content area. Similarly, descriptors that related to aspects of written responses and problem solving were presented together under the more general descriptor that asked whether the item called for the more general skill. For example, in the 8th-grade instrument, the written-response descriptors were grouped as follows:

"If the item requires a written response, check any of the following descriptions that apply:

	The item calls for:
22(a)	making conjectures
22(b)	defending ideas
22(c)	giving supporting examples
22(d)	explaining the reasoning process underlying conclusions
22(e)	conveying underlying reasoning skills beyond the level of arithmetic."

A number of versions of the instruments were piloted and revised before arriving at the final versions which are in Appendix C.[2] The final version of the instruments maintained the exact language of the NAGB achievement-level descriptions, unless there were semantic difficulties in leaving parsed clauses intact but separate.[3] When a clause had the connector "and" (depicting intersection of knowledge and skill types), it was typically switched to "or" so that an item requiring either knowledge or skill would be mapped to that descriptor.

The final instruments covered the knowledge and skills mentioned in the NAGB descriptions nearly completely. The attributes that were not included in the mapping protocol were of several specific types. One exception was references to the use of calculators, rulers and geometric shapes.[4] A second category of omissions were phrases that could not be viewed as a characteristic of a single item. For example, a number of phrases referred to demonstrating a skill "...in the five NAEP content areas." Finally, a few phrases referred to qualities of student performance rather than to skills or knowledge; for example, that students should be able to "use ... appropriately" or "display mastery in the use of ...."

Table 1 indicates the number of descriptors derived from each achievement level description at each grade.

Mapping Items to Elements of the Achievement Level Descriptions

Sample of judges. For each grade, a group of six mathematics educators (teachers or former teachers), who were familiar with the content of the mathematics curriculum at that grade, were recruited and trained to examine each test item and select the descriptors that described the knowledge or skills that "the item called for." A copy of the questionnaire used to gather

---

[2] In the final instruments, a question mark was placed after each descriptor so that judges could indicate uncertainty in their mapping of a particular descriptor to an item. However, there appears to be no systematic benefit from taking the reported uncertainty data into account.

[3] Some verbs were converted to gerunds (for example, "apply" became "applying").

[4] The items that require students to use calculators are grouped in particular blocks, and it is obvious which items call for the use of calculators; it is also obvious which items require use of geometric shapes and rulers.

background information on the judges and a summary of the background characteristics of the judges are presented in Appendix D.[5]

Data collection. To make the data collection process more efficient, the descriptor mapping protocol was computerized. HyperCard stacks were created to correspond to the total number of test items to be judged at each grade level. For each test item, a judge could select the descriptors that mapped the item by placing the cursor on the appropriate box and clicking the mouse button. Each judge's selections for each item were recorded and later compiled into one of three data files (one file for each grade level). A copy of the computer screens for the Grade 8 protocol appears in Appendix E.

Each judge received a binder that contained 14 blocks of 1992 NAEP mathematics test items at one grade level. These binders contain all the item blocks administered to the main NAEP sample; blocks used only for trend analysis purposes or other special studies were not included. Judges working at the 4th-grade level had 178 items to judge; judges working at the 8th-grade level considered 211 items; and judges working at the 12th-grade level covered 208 items.[6] At each grade level, half of the judges received the blocks of items in reverse order.

Judges were told that they were participating in a study whose purpose was to determine the mathematics knowledge and skills that are being

Table 1

Number of Descriptors Abstracted From NAGB Descriptions by Achievement Level and Grade

	Achievement level			
Grade	Basic	Proficient	Advanced	Total
4	5	9	4	18
8	8	17	6	31
12	14	14	7	35

[5] The summary of the teacher background questionnaire data was prepared by Audrey McEvans.

[6] The item pools for 4th, 8th, and 12th grades were not independent. Eighty-nine items were administered at both 4th and 8th grade; 92 items were administered at both 8th and 12 grade; and 34 items were administered at all three grade levels.

assessed by the NAEP mathematics test items. They were asked to **"use their own professional judgment in deciding which descriptors applied to each item and to interpret the descriptors in light of their experience of 4th-, 8th- or 12th-grade mathematics content and students."** The judges were told that there were no right or wrong decisions regarding which descriptors mapped to any item, and they were encouraged to select as many of the descriptors as applied to each item. However, to ensure independence in judgment, they were told not to discuss the descriptors or test items with any other judge. A copy of the script used to train the judges is contained in Appendix F. Judges were not given any information about either the existence of achievement levels or the identity of the achievement level from which each descriptor was taken prior to completing the mapping protocol.

The task of mapping items to descriptors took the judges an average of seven hours to complete. On completion of the task, the judges were given the text of the original NAGB achievement level descriptors. Then, all judges were asked to write their impressions of the activity and approximately half of them were interviewed. A report of the judges' comments is presented in Appendix G.[7]

The mapping of items to descriptors produced a considerable amount of information from each judge and across judges. Essentially, the six judges at Grade 4 each made 3204 (178 items x 18 descriptors) decisions mapping items to descriptors. The corresponding numbers for decisions at Grades 8 and 12 were 6541 (211 x 31) and 7280 (208 x 35), respectively.

Determining when items are mapped to descriptors and levels. A critical decision was what constituted a map between an item and a descriptor; that is, how to determine whether an item mapped to a descriptor and, through the descriptor's location in NAGB's achievement level descriptions, to an achievement level. We considered several possible decision rules (requiring that at least 4, 5, or all 6 judges map the item to the descriptor) and examined their empirical consequences. The more stringent criteria of complete (6) or almost complete (5) agreement would have resulted in too few items being mapped to any descriptors. If the criterion of complete agreement among all six judges had been selected, 132 4th-grade (74%), 116 8th-grade (55%), and 177

[7] Interviews were conducted by Regie Stites who also observed and prepared the summary of the judges' comments which appears in Appendix G.

12th-grade (85%) items would not have been mapped to any descriptors. If the criterion of at least 5 judges had been selected, 81 4th-grade items (46%), 30 8th-grade items (14%), and 120 12th-grade items (58%) would not have been matched to any descriptor. Applying the criterion that four judges had to agree on a mapping, there were 28 (out of 178) 4th-grade items, only 2 (out of 211) 8th-grade items, and 34 (out of 208) 12th-grade items that were not mapped to any descriptor. In the end, the criterion that **at least four of the six judges assign the descriptor to the item to consider it a mapping** was chosen.

With the chosen decision rule on item-descriptor mapping, each item was initially classified as representing an achievement level if at least four of the six judges assigned at least one descriptor from the particular achievement level to the item. Thus, an item could be assigned to more than one achievement level or, indeed, to no achievement level if there was no descriptor that was assigned to it by at least four judges.

To identify sets of items for each of the three achievement levels, three alternative rules were considered for assigning items that had been mapped to descriptors from multiple levels:

1. **Highest Level**. Each item with multiple level classifications was assigned to the **highest achievement level** from which even one descriptor was mapped to the item by four or more judges. This approach assumes that if an item calls for multiple skills, then it is the most advanced of those skills that limits performance on the item;

2. **Predominant Level**. Each item was assigned to the level from which the **majority of** the **descriptors** mapped to the item came; if there were an equal number of descriptors from two or three levels mapped to the item, the item was assigned to all of those levels;

3. **All Levels Where Mapped**. Each item was assigned to **all of the levels from which a descriptor was mapped to the item.**

Only the first rule resulted in a unique level assignment for each item; therefore the sets of items generated by this approach were used to examine the extent to which student performance statistics reflected the achievement level descriptions.[8]

---

[8] The decision to assign an item to the highest level from which a descriptor was mapped to the item is supported by the fact that almost all of the median *p*-values on these sets of items were less than .65 for students in all but the highest level from which a descriptor was assigned. See tables in Appendix K.

Summarizing Performance on Sets of Items

The performance of students scoring in the range of each achievement level was obtained for the sets of items assigned to individual descriptors, and multiple and single achievement levels. Specifically, the median *p*-values (percentages of students answering 50% of the items correctly) for items in a set were obtained for students whose NAEP scores fell in the regions classified as Below Basic, Basic, Proficient, or Advanced.

Different statistical criteria might be chosen to judge whether the performance of students on items mapped to descriptors at a given achievement level was consistent with the level from which the descriptor was abstracted. Reckase (1993) defined proficiency at a particular level as answering correctly a high proportion of items randomly sampled from that domain (or level or descriptor). In addition, in our view, the proportion of items answered correctly by students scoring below the level associated with the descriptor should be substantially lower than that for the level with which the descriptor is associated. Therefore, two criteria were chosen here for determining if the pattern of student performance is consistent with the descriptions of what students at each level should be able to do: The median *p*-values on the subset of items to which the descriptor was mapped should be at least .65 for students classified at the achievement level from which the descriptor was abstracted, and the *p*-values should be less than .5 for students classified at the next lowest level.[9] In other words, at least 65% of students scoring at the achievement level represented by a set of items should give the correct answer to half of the items in that set, and at least 50% of students scoring in the next lowest achievement level should get at least half those items wrong.

[9] We started out with an additional third criterion for deciding if the pattern of *p*-values reflected the level to which an item was classified. That third criterion was that there should be at least a .3 difference between the *p*-value for the target level (at least .65) and the *p*-value for the next lower level (less than .5). However, as the analysis progressed, we relaxed that criterion because using only the other two criteria resulted in a clearer and more easily interpreted pattern of results.

Results

The presentation of results of the study will be organized around the three questions addressed by the study. Those questions related to (a) the clarity of the achievement level descriptions; (b) the potential of the 1992 item pool to provide information related to the skills included in the descriptions; and (c) the performance of students on sets of items that call for the skills included in the achievement level descriptions.

Clarity of the Achievement Level Descriptions

The lists of distinct skills that were abstracted and presented to the judges were designed to increase the reliability of interpretations by focusing judges' attention on particular skills. However, because the language of the achievement level descriptions was preserved in the descriptors, the ambiguity inherent in the terminology of NAGB's descriptions and the similarity of phrases used in descriptions of different levels were also preserved. Table 2 contains several instances of descriptors from different levels with similar wording.

Aspects of the judges' mapping of descriptors to items and the written and oral comments of the judges provide indications of the lack of clarity of the achievement level descriptions. Judges (in posttask interviews and written comments) reported difficulty in deciding if a descriptor applied to an item when the descriptor was ambiguous or when there were multiple descriptors containing similar phrases. One judge reported, "Solving a 'simple real-world problem' versus solving a '[routine] real-world problem' drove me wild—purely a semantics problem. What is the difference? What is real-world?" Another judge wrote, "[M]any of the descriptors were very ambiguous and subjective—'unique,' 'complex,' 'fundamental,' 'mathematical ideas' are open to interpretation." Yet another wrote: "I tried to be consistent with the descriptors as they applied to different questions [test items] but I'm afraid I did not always accomplish [the] task.... I felt that some descriptors were intentionally ambiguous." Many judges indicated that they would have liked the opportunity to discuss the interpretation of the descriptors with other judges.

Table 2

Descriptors With Similar Phrases From Different Levels

Level	Descriptor ID number	Phrase
Grade 4		
B	D1a	(use basic facts to perform simple) computations with whole numbers
P	D1c	(use) whole numbers to compute results
B	D1b	estimate with whole numbers
P	D1d	(use) whole numbers to estimate
B	D2a	(show some) understanding of fractions and decimals
P	D2b	(have a conceptual) understanding of fractions and decimals
P	D8b	explanations of how solutions were achieved
A	D8c	explaining (why, as well as) how, answers (and solution processes) were achieved
B	D3	understanding (the mathematical) concepts and procedures
P	D4	procedural and conceptual understanding (to problem solving)
A	D5	procedural and conceptual understanding (to complex and non routine real-world problem solving)
B	D6a	solve (some simple) real-world problems
P	D6b	solve real-world problems
A	D6c	solve (complex and non routine) real-world problems
Grade 8		
B	D1	understanding of arithmetic operations
P	D2	(thorough) understanding of (basic-level) arithmetic operations
B	D15	conceptual and procedural (understanding)
P	D16	(applying mathematical) concepts and procedures
P	D22e	(convey) underlying reasoning (skills)
A	D22d	(explain) the reasoning (process) underlying their conclusion

Table 2 (continued)

Level	Descriptor ID number	Phrase
Grade 12		
B	D1a	(using) geometric reasoning (strategies)
P	D1c	(an understanding of) geometric reasoning
B	D2a	(using) algebraic reasoning
P	D2b	(an understanding of) algebraic reasoning
B	D9	reasonableness of results as applied to real-world (problems)
P	D10	reasonableness of answers as applied to real-world (situations)
B	D4a,b	(apply) statistical reasoning
P	D4c	(an understanding of) statistical reasoning
B	D14b	use (correct) mathematical (language) and symbols to communicate mathematical reasoning (processes)
A	D14c	communicate their mathematical reasoning through (clear, concise, and correct) use of mathematical symbolism
P	D3a	understand (elements of) the function concept
A	D3b	understand the function concept
B	D11	procedural and conceptual knowledge
P	D12	mathematical concepts and procedures
A	D13	procedural and conceptual knowledge

Tables 3, 4, and 5 display the descriptors that resulted in more than 9 cases (items) where the judges were evenly divided in applying the descriptor.[10] Descriptors with many evenly-divided decisions can be viewed as the descriptors that are least clear. In Grade 4, descriptors referencing

[10] At Grade 4, descriptor D3 (understanding of mathematical concepts or mathematical procedures) and at Grade 8, descriptor D15 (conceptual understanding or procedural understanding) were excluded because they were mapped to almost every item by at least four judges. In retrospect, the decision to leave mathematics concepts and mathematics procedures combined and conceptual and procedural understanding combined in a single descriptor was an unfortunate decision. Judges rightly concluded that virtually every item at these grades involved either conceptual or procedural understanding and responded accordingly. As a consequence these descriptors could not inform the mapping of items to achievement levels and thus were excluded.

Table 3

Descriptors for Which There Were More Than 9 Items With Evenly-Divided Judgments, Grade 4

Descriptor ID number	Descriptor text	Achievement level	# of items
D1a	using basic number facts to perform simple computations with whole numbers	B	39
D2a	some understanding of fractions or decimals	B	14
D6a	solving a simple real-world problem	B	13
D1b	estimating with whole numbers	B	10
D7	employing problem-solving strategies such as identifying and using appropriate information	P	62
D4	applying integrated procedural and conceptual understanding to problem solving	P	39
D1e	determining of the reasonableness of whole number results	P	26
D1c	using whole numbers to compute results	P	20
D6b	solving a [routine] real-world problem	P	18
D8d	clear or concise communication	A	12

whole numbers, problem solving, or clear and concise communication were least consistently interpreted. In Grade 8, descriptors referencing problem solving, reasonableness of answers and the general (gray-shaded) sentence for the Basic level, which contained references to multiple content and skills, were the least consistently interpreted. In Grade 12, descriptors referencing problem solving and understanding resulted in most evenly-divided decisions. Fewer Advanced descriptors than Basic or Proficient descriptors resulted in evenly-divided interpretations.

The results of a series of large-scale generalizability analyses of the mappings of descriptors to the items by the judges (reported in Novak, Burstein, and Sugrue, forthcoming) confirm that large sources of variability in mapping items to descriptors, especially at Grades 4 and 8, were descriptors and interactions of judges with descriptors. There was considerable variability in judges' interpretations of some clusters of descriptors; in general, the

Table 4

Descriptors for Which There Were More Than 9 Items With Evenly-Divided Judgments, Grade 8

Descriptor ID number	Descriptor text	Achievement level	# of items
D1	an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions or percents	B	55
D9	solving problems through the appropriate selection and use of strategies	B	51
D8	completing problems with the help of structural prompts such as diagrams, charts, or graphs	B	25
D10	solving problems through the appropriate selection and use of technological tools—including calculators, computers, or geometric shapes	B	21
D6a	using informal geometric concepts in problem solving	B	16
D12	determining which of available data are necessary and sufficient for correct solutions	B	16
D15	conceptual understanding or procedural understanding	B	15
D4	using fundamental algebraic concepts in problem solving	B	12
D7	familiarity with quantity or spatial relationships in problem solving or reasoning	P	39
D2	a thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical situations	P	18
D16	applying mathematical concepts and procedures to complex problems	P	12
D3	understanding the connections among any of the following: fractions, percents, decimals	P	11
D6b	applying the properties of informal geometry	P	11
D13a	making of inferences from data or graphs	P	10
D21	using number sense to consider the reasonableness of an answer	A	45
D6c	using geometric awareness to consider the reasonableness of an answer	A	23
D11	using abstract thinking to create unique problem-solving techniques	A	11

Table 5

Descriptors for Which There Were More Than 9 Items With Evenly-Divided Judgments, Grade 12

Descriptor ID number	Descriptor text	Achievement level	# of items
D11	procedural knowledge or conceptual knowledge in solving problems	B	57
D6	recognizing relationships presented in verbal, algebraic, tabular, or graphical forms	B	30
D2a	using algebraic reasoning strategies to solve problems	B	23
D1a	using geometric reasoning strategies to solve problems	B	21
D14a	using mathematical language and symbols to communicate mathematical relationships using mathematical language and symbols to communicate reasoning processes	B	14
D1b	knowledge of geometric relationships and corresponding measurement skills	B	13
D8	using estimation to verify solutions to real-world problems	B	13
D4e	generalizing from patterns or examples	B	10
D2b	an understanding of algebraic reasoning	P	18
D10	judging or defending the reasonableness of answers as applied to real-world situations	P	14
D1c	an understanding of geometric reasoning	P	13

descriptors that were least consistently mapped to items were those that did not reference specific mathematics content.[11]

It is not surprising to find less consistent interpretation of process than content aspects of the descriptors. In general, it is easier to focus on the

[11] Further evidence for a lack of distinction between descriptors from different levels comes from analysis of the descriptors that were mapped to the 34 items that were the same across all three grades. One would expect that the descriptors mapped to these items would belong to similar or lower achievement levels as the grade level increases. This was not always the case, particularly when the 4th- and 8th-grade mappings were compared. The 8th-grade judges mapped 20 items to descriptors from higher levels than the items had been mapped to by the 4th-grade judges. Tables showing the number of multi-grade items assigned to each level across pairs of grades are presented in Appendix L.

content targeted by a test item and easier to validly interpret performance in relation to content domains than to process domains such as "problem solving" or "understanding." However, it is often more useful to know what level of understanding a student has or what kind of problems a student can solve within a content domain, than to have just a score that represents some general ability in that domain. The cognitive demands of the test items used to generate the overall score are critical to interpreting that score. It is difficult to estimate the cognitive demands of a test item without analysis of student responses to the item.

Written and oral comments from the judges indicate that not only ambiguity of the descriptors but also features of the test items were a source of difficulty in making judgments. One judge referred to "implicit" and "explicit" features of items, the implicit features being less easy to define in terms of the descriptors. Another judge said, "[in] classifying some items, a lot would depend on the type of test taker a person is—many people estimate all answers rather than working them out, or generate when they are not require to." The problem of multiple possible approaches to answering a test item inevitably arises when one attempts to define the cognitive demands of an assessment task (French, 1965; Linn, Baker, & Dunbar, 1991). However, in this study, judges were allowed to apply multiple descriptors to any one item, thereby permitting identification of multiple possible strategies for solving the same problem. Indeed, a number of judges indicated that they tried to imagine multiple possible solutions to items.

Descriptors that were consistently mapped to a large number of items, but also had a large number of evenly-divided judgments, might indicate cases where the inconsistency of judges' mapping is more a function of the items than the descriptors. Tables 6, 7, and 8 display such descriptors; that is, descriptors that were consistently mapped to more than 9 items but were also evenly divided for more than 9 items. Very few descriptors that were mapped to more than 9 items by at least four judges did not also receive an equally large number of evenly-divided judgments.

For example, the fourth-grade Proficient descriptor D7 was consistently mapped to 70 items, but had evenly-divided decisions on 62 items. Fourth grade descriptor D7 asked judges to decide if an item called for "employing

Table 6

Descriptors With Consistent Mappings to More Than 9 Items and Evenly-Divided Judgments for More Than 9 Items, Grade 4

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)	# of items (3 judges)
D1b	estimating with whole numbers	B	14	10
D1a	using basic number facts to perform simple computations with whole numbers	B	10	39
D4	applying integrated procedural and conceptual understanding to problem solving	P	87	39
D7	employing problem-solving strategies such as identifying and using appropriate information	P	70	62
D1c	using whole numbers to compute results	P	56	20
D6b	solving a [routine] real-world problem	P	46	18
D1e	determining of the reasonableness of whole number results	P	10	26
D8d	clear or concise communication	A	11	12

problem-solving strategies such as identifying and using appropriate information." Therefore one might conclude that for many items, it was obvious that the item called for strategies such as identifying and using appropriate information, but there was also a large number of items where it was less obvious, and so judges differed in their impression of whether or not those items called for those skills.

In summary, the variability of decisions made by judges in mapping some descriptors to items indicates that the achievement level descriptions lack clarity. Many of the skills included in the achievement level descriptions are too ambiguous to engender common interpretation, and there is not enough distinction between the description of skills associated with separate levels. In addition, there is inconsistency between the general (shaded) sentences and the skills described in the ensuing paragraphs. There is nonsystematic mention of specific content areas in addition to "the five content areas," and there is also nonsystematic mention of skills that students scoring at a particular level should NOT possess.

Table 7

Descriptors With Consistent Mappings to More Than 9 Items and Evenly-Divided Judgments for More Than 9 Items, Grade 8

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)	# of items (3 judges)
D9	solving problems through the appropriate selection and use of strategies	B	89	51
D8	completing problems with the help of structural prompts such as diagrams, charts, or graphs	B	81	25
D1	an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions or percents	B	78	55
D6a	using informal geometric concepts in problem solving	B	48	16
D10	solving problems through the appropriate selection and use of technological tools—including calculators, computers, or geometric shapes	B	42	21
D4	using fundamental algebraic concepts in problem solving	B	15	12
D2	a thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical situations	P	108	18
D7	familiarity with quantity or spatial relationships in problem solving or reasoning	P	77	39
D6b	applying the properties of informal geometry	P	49	11
D3	understanding the connections among any of the following: fractions, percents, decimals	P	10	11
D21	using number sense to consider the reasonableness of an answer	A	50	45
D6c	using geometric awareness to consider the reasonableness of an answer	A	16	23

Distribution of Items Across Descriptors

In spite of the variability of judges' application of some descriptors, there were enough consistent mappings (by at least four judges) to warrant analysis of the extent of coverage by the 1992 mathematics assessment of skills depicted

Table 8

Descriptors With Consistent Mappings to More Than 9 Items and Evenly-Divided Judgments for More Than 9 Items, Grade 12

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)	# of items (3 judges)
D11	procedural knowledge or conceptual knowledge in solving problems	B	53	57
D1b	knowledge of geometric relationships and corresponding measurement skills	B	46	13
D2a	using algebraic reasoning strategies to solve problems	B	36	23
D1a	using geometric reasoning strategies to solve problems	B	22	21
D6	recognizing relationships presented in verbal, algebraic, tabular, or graphical forms	B	19	30
D2b	an understanding of algebraic reasoning	P	17	18
D1c	an understanding of geometric reasoning	P	12	13

in the achievement level descriptions. Tables 9, 10, and 11 show the number of items that were mapped to each descriptor by at least four judges. The descriptors are ordered by achievement level and by the number of items mapped to the descriptor by at least four judges because that was the criterion finally chosen to decide if an item should be considered appropriate for measuring the knowledge or skill implied by the particular descriptor. Applying the criterion that four judges had to agree on a mapping, there were 158 (out of 178) 4th-grade items, 209 (out of 211) 8th-grade items, and 174 (out of 208) 12th-grade items that were mapped to at least one descriptor.

There is great variation in the distribution of items across descriptors. Some descriptors were mapped (by at least four judges) to very few items. For example, at Grade 4, 7 of the 18 descriptors were mapped to fewer than 9 items; the same was true for 16 of 31 Grade 8 descriptors and 24 of 35 Grade 12 descriptors. Therefore, in both Grades 8 and 12, more than half of the skills included in the achievement level descriptions were mapped to less than 9 items in the 1992 item pool.

Table 9

Number of Items Mapped to Each Descriptor by at Least 4 Out of 6 Judges, Grade 4

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)
D1b	estimating with whole numbers	B	14
D1a	using basic number facts to perform simple computations with whole numbers	B	10
D2a	some understanding of fractions or decimals	B	4
D6a	solving a simple real-world problem	B	1
D4	applying integrated procedural and conceptual understanding to problem solving	P	87
D7	employing problem-solving strategies such as identifying and using appropriate information	P	70
D1c	using whole numbers to compute results	P	56
D6b	solving a [routine] real-world problem	P	46
D2b	conceptual understanding of fractions or decimals	P	13
D1d	using whole numbers to estimate results	P	12
D1e	determining of the reasonableness of whole number results	P	10
D8b	explaining how the answer or solution process was achieved	P	8
D8a	giving supporting information	P	6
D8d	clear or concise communication	A	11
D8c	explaining why the answer or solution process was achieved	A	3
D5	applying integrated procedural and conceptual understanding to complex and nonroutine real-world problem solving	A	1
D6c	solving a complex and nonroutine real-world problem	A	1

Table 10

Number of Items Mapped to Each Descriptor by at Least 4 Out of 6 Judges, Grade 8

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)
D9	solving problems through the appropriate selection and use of strategies	B	89
D8	completing problems with the help of structural prompts such as diagrams, charts, or graphs	B	81
D1	an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions or percents	B	78
D6a	using informal geometric concepts in problem solving	B	48
D10	solving problems through the appropriate selection and use of technological tools—including calculators, computers, or geometric shapes	B	42
D4	using fundamental algebraic concepts in problem solving	B	15
D12	determining which of available data are necessary and sufficient for correct solutions	B	2
D2	a thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical situations	P	108
D7	familiarity with quantity or spatial relationships in problem solving or reasoning	P	77
D6b	applying the properties of informal geometry	P	49
D14a	calculating results within the domain of statistics or probability	P	13
D3	understanding the connections among any of the following: fractions, percents, decimals	P	10
D13a	making of inferences from data or graphs	P	9
D22e	conveying underlying reasoning skills beyond the level of arithmetic	P	9
D22b	defending ideas	P	8
D22c	giving supporting examples	P	6
D19	generating one's own examples	P	6

Table 10 (continued)

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)
D14b	evaluating results within the domain of statistics or probability	P	6
D16	applying mathematical concepts and procedures to complex problems	P	6
D5	understanding of the connection between algebra and functions	P	3
D13b	understanding of the process of gathering and organizing data	P	3
D22a	making conjectures	P	2
D14c	communicating results within the domain of statistics or probability	P	2
D18	comparing and contrasting mathematical ideas	P	1
D21	using number sense to consider the reasonableness of an answer	A	50
D6c	using geometric awareness to consider the reasonableness of an answer	A	16
D22d	explaining the reasoning process underlying conclusions	A	15
D11	using abstract thinking to create unique problem-solving techniques	A	2
D17	reaching beyond the recognition, identification, and application of mathematical rules to generalize and synthesize concepts and principles	A	1
D20	probing of examples and counter examples in order to shape generalizations from which the student can develop models	A	1

Table 11

Number of Items Mapped to Each Descriptor by at Least 4 Out of 6 Judges, Grade 12

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)
D11	procedural knowledge or conceptual knowledge in solving problems	B	53
D1b	knowledge of geometric relationships and corresponding measurement skills	B	46
D2a	using algebraic reasoning strategies to solve problems	B	36
D1a	using geometric reasoning strategies to solve problems	B	22
D6	recognizing relationships presented in verbal, algebraic, tabular, or graphical forms	B	19
D4b	applying statistical reasoning in reading tables or graphs	B	9
D14b	using mathematical language and symbols to communicate reasoning processes	B	8
D14a	using mathematical language and symbols to communicate mathematical relationships	B	8
D9	using estimation to determine the reasonableness of results as applied to real-world problems	B	7
D2d	generalizing from patterns or examples	B	5
D8	using estimation to verify solutions to real-world problems	B	5
D4e	generalizing from patterns or examples	B	2
D1f	generalizing from patterns or examples	B	1
D4a	applying statistical reasoning in the organization and display of data	B	1
D4d	analyzing and interpreting data in tabular or graphical form	P	18
D2b	an understanding of algebraic reasoning	P	17
D14d	defending ideas	P	16
D1d	an understanding of spatial reasoning	P	13

Table 11 (continued)

Descriptor ID number	Descriptor text	Level	# of items (4 or more judges)
D1c	an understanding of geometric reasoning	P	12
D1e	justifying geometric relationships	P	10
D2c	performing algebraic operations involving polynomials	P	9
D14f	giving supporting examples	P	8
D4c	an understanding of statistical reasoning	P	7
D10	judging or defending the reasonableness of answers as applied to real-world situations	P	7
D3a	understanding of elements of the function concept in symbolic, graphical or tabular form	P	3
D14e	making conjectures	P	1
D3c	using elements of the function concept in symbolic, graphical or tabular form	P	1
D12	integrating mathematical concepts and procedures to the solution of more complex problems	P	0
D14c	clear and concise use of mathematical symbolism and logical thinking to communicate mathematical reasoning	A	5
D3e	applying the numeric, algebraic, or graphical properties of functions	A	4
D3b	understanding of the function concept	A	2
D3d	comparing the numeric, algebraic, or graphical properties of functions	A	1
D5	solution of problems in the more advanced area of continuous and discrete mathematics	A	1
D13	the integration of procedural and conceptual knowledge, and the synthesis of ideas	A	1
D7	formulating generalizations and creating models through probing examples and counterexamples	A	0

Even though there were many more descriptors at Grade 12 than at Grade 4, there were only a few 12th-grade descriptors that were consistently mapped to a large number of items. The 12th-grade descriptors that were consistently mapped to more than 9 items involved straightforward topic/content terms (geometric relationships and corresponding measurement skills, algebraic reasoning strategies, reading tables and graphs, analyzing and interpreting data in tabular or graphical form) rather than more ambiguous references to cognitive processes. The 12th-grade Advanced descriptor involving explicit requests to defend one's ideas in written responses was also consistently mapped.

Distribution of Items Across Levels

Items were assigned to achievement levels based on the level of the descriptors that were mapped to each item by at least four judges. The number of items assigned to descriptors from each achievement level by at least 4 judges is reported in Table 12. Since each item could be mapped to more than one descriptor, a considerable number of items were mapped to descriptors from multiple levels.[12] This was the case particularly in Grades 8 and 12; over half the Grade 8 items and almost a quarter of the 12th-grade items were mapped to descriptors at both Basic and Proficient levels. The numbers of descriptors from each achievement level description that were consistently mapped to each item are presented in Appendix H.

Three different rules were used to assign to levels the items that were mapped to descriptors from more than one level. The first rule assigned these items to the highest achievement level from which even one descriptor was mapped to the item by four or more judges. This led to the assignment of all of the 8th- and 12th-grade items that were mapped to descriptors from both the Basic and Proficient levels to the Proficient level; all 58 8th-grade items that were mapped to descriptors from all three levels were assigned to the

[12] Since judges could match more than one descriptor to any one item, there was variation in the number of descriptors that were mapped to different items. The number of descriptors mapped to any one item ranged from none to 6 for 4th-grade items, none to 11 for 8th-grade items, and none to 8 for 12th grade items. Fifty-six percent of 4th-grade items, 91% of 8th-grade items, and 48% of 12th-grade items were mapped to more than one descriptor. However, only 8 of the 178 8th-grade items and 17 of the 208 12th-grade items were mapped to more than 4 descriptors; 56 of the 211 8th-grade items were mapped to more than 4 descriptors.

Table 12

Number of Items Classified to Single or Multiple Achievement Levels

Level	Grade 4	Grade 8	Grade 12
Not Classified	28	2	34
Basic	6	13	88
Proficient	109	11	23
Basic & Proficient	22	110	50
Advanced	1	0	2
Basic & Advanced	0	10	1
Proficient & Advanced	12	7	3
Basic & Proficient & Advanced	0	58	7
Total	178	211	208

Note. Items were classified to a given level if at least 4 out of 6 judges mapped at least one descriptor from the particular achievement level to the item.

Advanced level. The resulting numbers of items assigned to single levels is presented in Table 13. Using this "highest level" rule, there appear to be very few items representative of the Basic level in Grades 4 and 8 or the Advanced level in Grades 4 and 12.

Table 13

Number of Items Classified to Highest Single Achievement Level

Level	Grade 4	Grade 8	Grade 12
Not Classified	28	2	34
Basic	6	13	88
Proficient	131	121	73
Advanced	13	75	13
Total	178	211	208

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

The second rule assigned each multilevel item to the level of the majority of the descriptors that were mapped to it, or, if an equal number of descriptors from two or three levels were mapped to the item, then the item would be assigned to all of those levels. Table 14 presents the numbers of items that were assigned to each level using the second, "predominant" level rule. Since items mapped to equal numbers of descriptors from more than one level were assigned to each of those levels, the totals in Table 14 exceed the actual number of 1992 items. The assignment of items based on the "predominant" rule still leaves very little coverage of 4th-grade Basic level skills and the Advanced level skills in all grades.

The third rule used to deal with the mapping of items to descriptors from different levels was to assign items to all levels from which a descriptor was mapped to the item. Table 15 presents the distribution of items to levels that result from application of the third or "all levels" rule. Sparse coverage of the Grade 4 Basic level, and the Advanced levels in Grades 4 and 12 are still indicated.

No matter which rule is used to assign items to levels, it appears that there is little coverage of skills associated only with the Basic level in Grade 4, and skills associated with the Advanced levels in Grades 4 and 12. For the remainder of this report, analyses of sets of items linked to levels will be reported for sets of items assigned to original single and multiple levels and for sets of items assigned to single unique levels based on the "highest level" rule.

Table 14

Number of Items Classified to Predominant Achievement Level

Level	Grade 4	Grade 8	Grade 12
Not Classified	28	2	34
Basic	18	159	133
Proficient	143	74	65
Advanced	3	21	6
Total	192	256	238

Note. Items were classified to the level from which most of the descriptors mapped to them were extracted.

Table 15

Number of Items Classified to "All" Single Achievement Levels

Level	Grade 4	Grade 8	Grade 12
Not Classified	28	2	34
Basic	28	191	145
Proficient	143	186	83
Advanced	13	75	13
Total	212	454	275

Note. Items were classified to all levels from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

Variation in distribution of items across levels by item format and content. Tables 16, 17, and 18 present the assignment of items to highest single levels by item format based on judges' mapping of descriptors to items. Tables showing the assignment of items of different formats to single and multiple levels are included in Appendix I.

At Grade 4, no constructed response or extended constructed response items were assigned to the Basic level; in fact at Grade 4, all five extended constructed response items were deemed to require at least one skill from the Advanced level. At Grade 8, more than half of the constructed response items were assigned to the Proficient level. At 12th grade, almost half of the constructed response items were judged to require skills included in the description of the Basic level. Therefore, as one advances through the grades, the range of achievement levels represented by constructed and extended constructed response items changes. There is also variation in the distribution of multiple-choice items across levels in different grades. At 4th grade, almost 80% of the multiple-choice items are classified as Proficient; at 8th grade, over 50% of the multiple-choice items are classified as Proficient and over 40% are classified as Advanced; at 12th grade, over 40% of items are classified as Basic and 36% are classified as Proficient, with less than 1% classified as Advanced.

Table 16

Number of Items Classified to Highest Single Achievement Level by Item Format, Grade 4

Level	Multiple choice	Constructed response	Extended constructed response	Total
Not classified	19	9	0	28
Basic	6	0	0	6
Proficient	93	38	0	131
Advanced	1	7	5	13
Total	119	54	5	178

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

Table 17

Number of Items Classified to Highest Single Achievement Level by Item Format, Grade 8

Level	Multiple choice	Constructed response	Extended constructed response	Total
Not Classified	1	1	0	2
Basic	9	4	0	13
Proficient	76	42	3	121
Advanced	60	12	3	75
Total	146	59	6	211

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

Table 18

Number of Items Classified to Highest Single Achievement Level by Item Format, Grade 12

Level	Multiple choice	Constructed response	Extended constructed response	Total
Not classified	25	9	0	34
Basic	63	24	1	88
Proficient	52	18	3	73
Advanced	5	6	2	13
Total	145	57	6	208

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

Tables 19, 20, and 21 present the assignment of items to highest single levels by content type (based on the NAEP assessment framework). Tables showing the assignment of items to multiple and single levels are included in Appendix I.

Table 19

Number of Items Classified to Highest Single Achievement Level by Item Content, Grade 4

Level	Numbers & operations	Measure-ment	Geometry	Data analysis	Algebra & functions	Estimation	Total
Not classified	12	2	11	2	1	0	28
Basic	1	0	0	0	0	5	6
Proficient	48	27	13	13	15	15	131
Advanced	2	2	3	5	1	0	13
Total	63	31	27	20	17	20	178

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

Table 20

Number of Items Classified to Highest Single Achievement Level by Item Content, Grade 8

Level	Numbers & operations	Measure-ment	Geometry	Data analysis	Algebra & functions	Estimation	Total
Not classified	2	0	0	0	0	0	2
Basic	6	3	0	2	2	0	13
Proficient	23	19	23	26	21	9	121
Advanced	27	10	13	6	6	13	75
Total	58	32	36	34	29	22	211

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

Table 21

Number of Items Classified to Highest Single Achievement Level by Item Content, Grade 12

Level	Numbers & operations	Measure-ment	Geometry	Data analysis	Algebra & functions	Estimation	Total
Not classified	11	4	2	6	4	7	34
Basic	24	18	15	1	18	12	88
Proficient	7	7	19	20	17	3	73
Advanced	2	0	1	2	8	0	13
Total	44	29	37	29	47	22	208

Note. Items were classified to the highest level from which at least one descriptor was mapped to the item by at least 4 out of 6 judges.

In Grade 4, a higher proportion of geometry items than items from the other content types were not classified to any level. This may be a reflection of the ambiguity of the process skills linked to geometry in the descriptions rather than to a lack of items related to geometry content. In Grade 8, a greater proportion of items of every content type except numbers and operations, and estimation, were assigned to the Proficient than to the Advanced level. At Grade 12, over half of the items assigned to the Advanced level were assessing

algebra and functions, but over 80% of the algebra and function items were assigned to either the Basic or Proficient levels. In addition, at Grade 12, most of the data analysis items were linked to descriptors from the Proficient level, and most of the estimation items that were classified were linked to the Basic level. Therefore, in addition to items being unevenly distributed across levels, items of some content types were judged to require skills associated with some level(s) more than other levels.

Analysis of Student Performance on Items Mapped to Descriptors and Levels

Performance on items mapped to individual descriptors. As was described earlier, a variant of the NAEP anchor item criteria was adopted as the criteria for judging whether the performance of students on a set of items mapped to a descriptor (by at least four out of six judges) was consistent with the level from which the descriptor was taken. Specifically, the median *p*-value on the subset of items mapped to a descriptor should be at least .65 for students scoring at the achievement level from which the descriptor was abstracted; in addition, the median *p*-value for students scoring in the next lowest level should be less than .5.[13] The median *p*-values across the set of items mapped to each descriptor are provided in Tables 22, 23, and 24 for all descriptors to which at least 9 items were mapped.[14]

---

[13] We started out with an additional third criterion for deciding if the pattern of *p*-values reflected the level to which an item was classified. That third criterion was that there should be at least a .3 difference between the *p*-value for the target level (at least .65) and the *p*-value for the next lower level (less than .5). However, as the analysis progressed, we relaxed that criterion because using only the other two criteria resulted in a clearer and more easily interpreted pattern of results.

[14] The mean, median, minimum and maximum *p*-value for sets of items mapped to each descriptor are presented in Appendix J.

Table 22

Median *P*-Values, for Students Classified at Each Level, on Sets of Items Mapped to Descriptors, Grade 4

Descriptor ID number	Descriptor text	Level of descriptor	# of items mapped to descriptors	Level of students			
				Below basic	Basic	Proficient	Advanced
D1a	using basic number facts to perform simple computations with whole numbers	B	10	.289	.472	.738	.933
D1b	estimating with whole numbers	B	14	.351	.468	.667	.898
D1c	using whole numbers to compute results	P	56	.264	.526	.792	.942
D1d	using whole numbers to estimate results	P	12	.287	.456	.706	.922
D1e	determining of the reasonableness of whole number results	P	10	.332	.496	.707	.900
D2b	conceptual understanding of fractions or decimals	P	13	.184	.275	.533	.898
D4	applying integrated procedural and conceptual understanding to problem solving	P	87	.208	.418	.660	.898
D6b	solving a [routine] real-world problem	P	46	.235	.449	.682	.922
D7	employing problem-solving strategies such as identifying and using appropriate information	P	70	.212	.410	.675	.890
D8d	clear or concise communication	A	11	.067	.228	.534	.680

Note. Only descriptors to which at least 9 items were mapped by 4 out of 6 judges are included.

Table 23

Median *P*-Values, for Students Classified at Each Level, on Sets of Items Mapped to Descriptors, Grade 8

Descriptor ID number	Descriptor text	Level of descriptor	# of Items mapped to descriptor	Level of students			
				Below basic	Basic	Proficient	Advanced
D1	an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions or percents	B	78	.343	.633	.829	.940
D4	using fundamental algebraic concepts in problem solving	B	15	.195	.373	.744	.956
D6a	using informal geometric concepts in problem solving	B	48	.248	.512	.710	.906
D8	completing problems with the help of structural prompts such as diagrams, charts, or graphs	B	81	.271	.541	.764	.930
D9	solving problems through the appropriate selection and use of strategies	B	89	.256	.523	.763	.930
D10	solving problems through the appropriate selection and use of technological tools—including calculators, computers, or geometric shapes	B	42	.248	.406	.685	.831
D2	a thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical situations	P	108	.265	.542	.811	.946
D3	understanding the connections among any of the following: fractions, percents, decimals	P	10	.205	.335	.685	.868

Table 23 (continued)

Descriptor ID number	Descriptor text	Level of descriptor	# of Items mapped to descriptor	Level of students			
				Below basic	Basic	Proficient	Advanced
D6b	applying the properties of informal geometry	P	49	.231	.482	.693	.904
D7	familiarity with quantity or spatial relationships in problem solving or reasoning	P	77	.358	.623	.845	.944
D13a	making of inferences from data or graphs	P	9	.381	.654	.804	.954
D14a	calculating results within the domain of statistics or probability	P	13	.179	.329	.721	.943
D22e	conveying underlying reasoning skills beyond the level of arithmetic	P	9	.004	.037	.160	.442
D6c	using geometric awareness to consider the reasonableness of an answer	A	16	.213	.287	.502	.740
D21	using number sense to consider the reasonableness of an answer	A	50	.459	.717	.896	.962
D22d	explaining the reasoning process underlying conclusions	A	15	.038	.232	.481	.762

Note. Only descriptors to which at least 9 items were mapped by 4 out of 6 judges are included.

Table 24

Median *P*-Values, for Students Classified at Each Level, on Sets of Items Mapped to Descriptors, Grade 12

Descriptor ID number	Descriptor text	Level of descriptor	# of Items mapped to descriptor	Level of students			
				Below basic	Basic	Proficient	Advanced
D1a	using geometric reasoning strategies to solve problems	B	22	.264	.361	.701	.892
D1b	knowledge of geometric relationships and corresponding measurement skills	B	46	.228	.354	.774	.940
D2a	using algebraic reasoning strategies to solve problems	B	36	.193	.356	.818	.947
D4b	applying statistical reasoning in reading tables or graphs	B	9	.329	.359	.647	.916
D6	recognizing relationships presented in verbal, algebraic, tabular, or graphical forms	B	19	.470	.679	.854	.949
D11	procedural knowledge or conceptual knowledge in solving problems	B	53	.340	.575	.868	.972
D1c	an understanding of geometric reasoning	P	12	.311	.530	.778	.859
D1d	an understanding of spatial reasoning	P	13	.325	.661	.864	.949
D1e	justifying geometric relationships	P	10	.437	.804	.898	.961
D2b	an understanding of algebraic reasoning	P	17	.153	.289	.714	.929
D2c	performing algebraic operations involving polynomials	P	9	.204	.498	.860	.971
D4d	analyzing and interpreting data in tabular or graphical form	P	18	.484	.726	.850	.936
D14d	defending ideas	P	16	.048	.170	.504	.690

Note. Only descriptors to which at least 9 items were mapped by 4 out of 6 judges are included.

Applying the ".65/.5" criteria for consistency of student performance with descriptor level,[15] it becomes evident that some of the descriptors would be more appropriate for describing the performance of students at a level other than the level from which the descriptor was abstracted. The 4th-grade results in Table 22 indicate that both Basic descriptors should be Proficient and one Proficient descriptor (D2b) should be Advanced. At Grade 8 (Table 23), five of the six Basic descriptors should be Proficient, two of the six Proficient descriptors should be Basic, and one Proficient descriptor (D22e) does not even have a median *p*-value of at least .65 for students classified as Advanced. One of the three advanced descriptors (D21) should be Basic. The results at Grade 12 (Table 24) are much the same. Five of the six Basic descriptors should be Proficient; three of the seven Proficient descriptors (D1d, D1e, and D4d) should be Basic; and there were no Advanced 12th-grade descriptors to which at least nine items were mapped.

Assuming that the assignment of items to levels on the basis of judges' decisions in this study is accurate, the pattern of performance reflected in Tables 22 to 24 raises questions about the appropriateness of some elements of NAGB's narrative descriptions of what students scoring at each level should be able to do. Of the 39 descriptors to which at least 9 items were mapped, less than half (17) exhibited a pattern of student performance that was consistent with the achievement level statements from which the descriptors were derived.

If one looks at the entire distributions for *p*-values of groups of students scoring at particular levels, across items mapped to particular descriptors, there are patterns that further call the achievement level descriptions into doubt. Illustrative distributions of the *p*-values across subsets of items mapped to four eighth-grade descriptors are displayed in Figures 1-4. These show that, within a set of items mapped to a descriptor, there is considerable variation in the percentage of students at a given level who answer different items correctly.

[15] In some cases we had to relax the less than .5 criterion slightly in our interpretations. Otherwise certain descriptors could not have been classified according to levels.

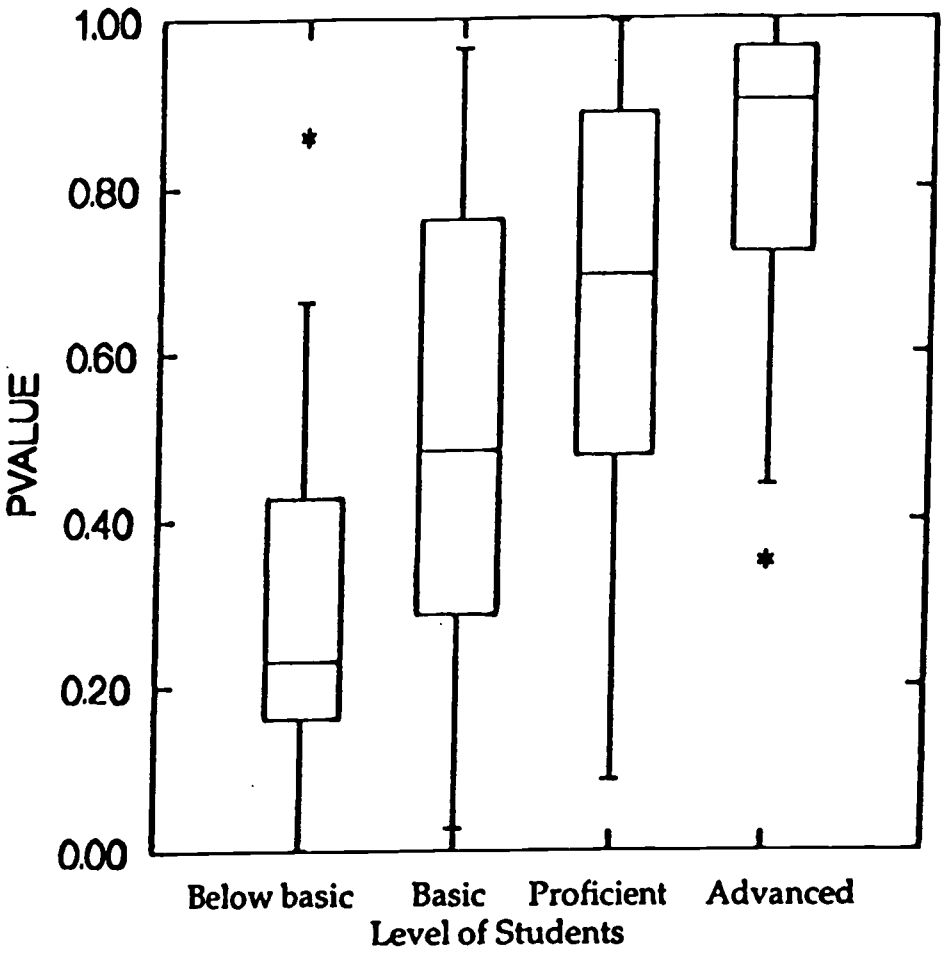


Figure 1. *P*-values for groups of students on set of 49 items mapped to Proficient descriptor number D6b (applying the properties of informal geometry), Grade 8.

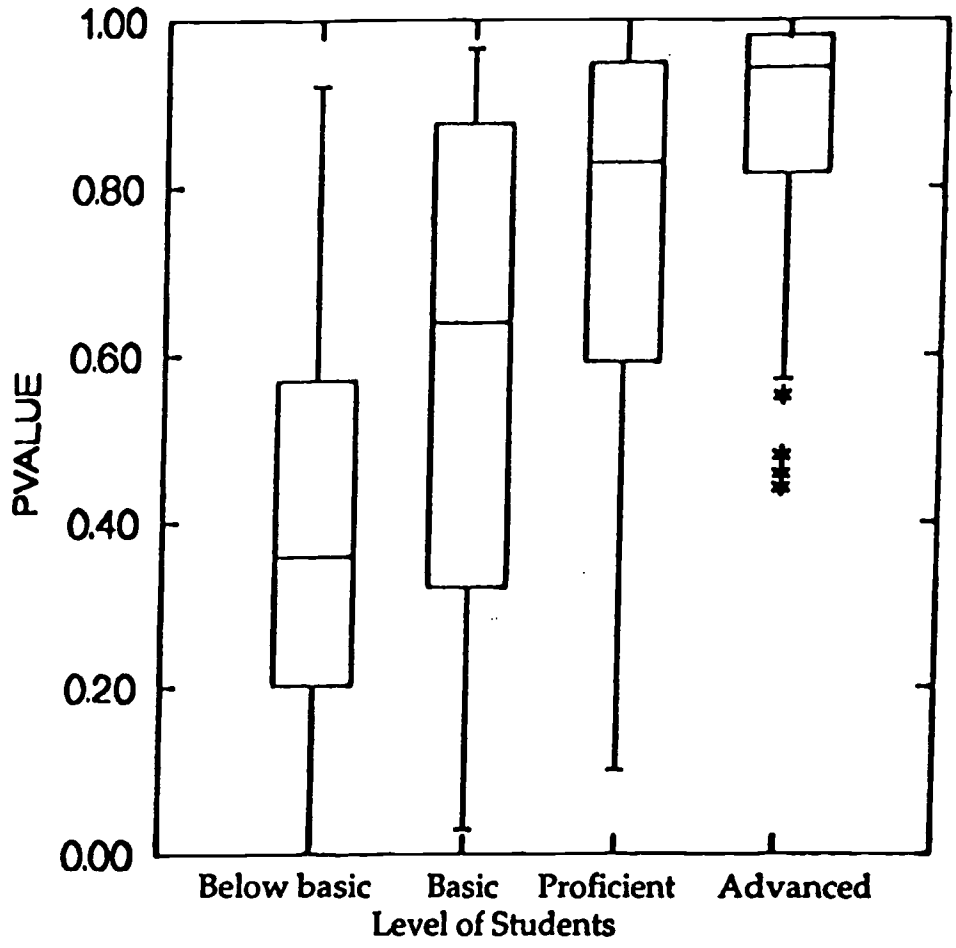


Figure 2. *P*-values for groups of students on set of 78 items mapped to Basic descriptor number D1 (an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, or percents), Grade 8.

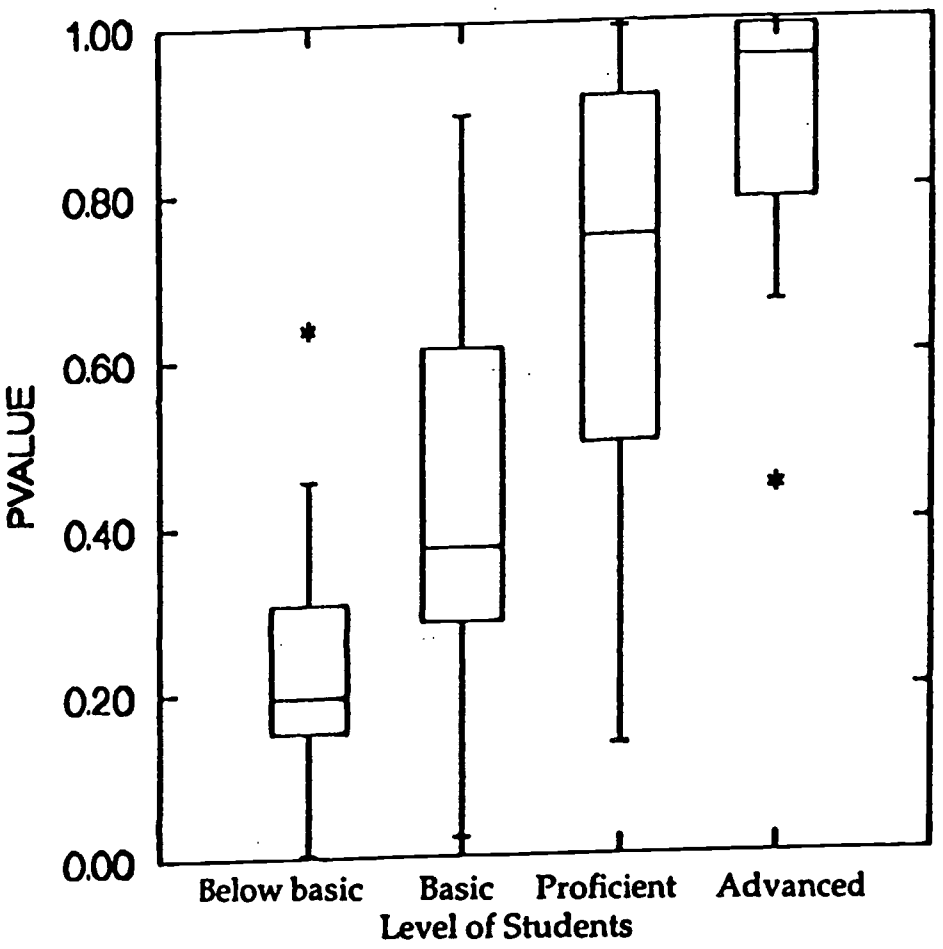


Figure 3. *P*-values for groups of students on set of 15 items mapped to Basic descriptor number D4 (using fundamental algebraic concepts in problem solving), Grade 8.

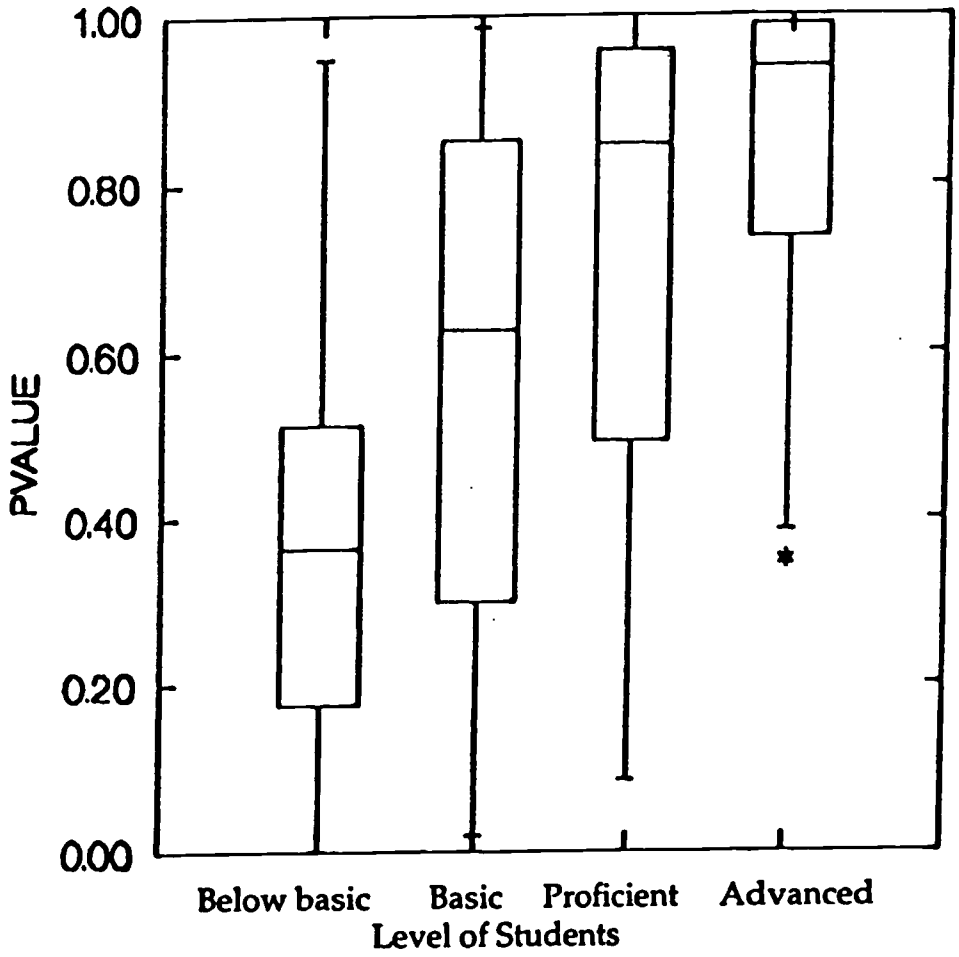


Figure 4. *P*-values for groups of students on set of 17 items mapped to Proficient descriptor number D7 (familiarity with quantity or spatial relationships in problem solving or reasoning), Grade 8.

The fact that the percent correct varies across items mapped to any given descriptor is in itself neither surprising nor undesirable. Reckase (1992) argues that, although the description of a skill such as those contained in the NAGB achievement level descriptions "defines a domain of items" (p. 1), there may be a very large number of items that match that description, and they vary in difficulty and discrimination over a fairly wide range. What is meant when someone says that students at the Proficient level can perform the necessary operations is that if students at that level were given a random sample of items from that domain, they would answer a high proportion of them correctly. However, it does not mean that they would be able to answer the hardest one correctly with high probability.

Variation in percent correct across items mapped to the same descriptor is not the only instance of variation in performance that is inconsistent with the achievement level descriptions. The distributions of *p*-values differ markedly across descriptors within the same achievement level (e.g., D1 versus D4); the distributions of *p*-values overlap considerably across the levels; and some of the distributions of *p*-values are very low. One would expect that for any item assigned to a descriptor, at least half of the students scoring at that level from which the descriptor was abstracted would answer the item correctly. That is not the case for many items.

If the pool of NAEP assessment items adequately represents the domains associated with specific descriptors (which it may not), Figures 1 to 4 serve to highlight what may be either misassignment of descriptor statements to achievement levels or simply flawed descriptions of the skills purportedly associated with certain levels. For example, the descriptor D4, "using fundamental algebraic concepts in problem solving," was drawn from the Basic level description; yet more than 75% of the 15 items mapped to this descriptor had percent correct values for Basic students less than the threshold of .65 (Figure 3). Conversely, the performance of the students scoring at the Basic level on items mapped to descriptor D7, "familiarity with quantity or spatial relationships in problem solving or reasoning," from the Proficient level description was distributed fairly evenly around .65 (Figure 4).

Performance on items mapped to levels. The lack of consistency between scores on the NAEP scale and performance on sets of items mapped to

elements of the achievement level descriptions is also evident if one considers performance on sets of items that were mapped to single achievement levels. Median *p*-values for groups of students scoring at each level on sets of items assigned uniquely to the highest single levels for Grades 4, 8 and 12 are presented in Tables 25, 26, and 27. (Tables showing the median *p*-values for sets of items assigned to single and multiple levels are included in Appendix K.) The median *p*-value for the set of items mapped to a particular level should be high (at least .65) for students classified at that level or higher, but lower (less than .5) for students classified at lower levels. This is the case, except for the sets of 4th- and 12th-grade Basic items. For these two sets of Basic items, the median *p*-value for Basic students was less than .65. In fact, a majority of the students who scored in the Basic range in Grade 4 got more than half of the Basic items wrong.

Table 25

Median *P*-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Highest Single Level, Grade 4

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	28	.318	.635	.829	.936
Basic	6	.429	.471	.673	.851
Proficient	131	.257	.475	.734	.924
Advanced	13	.067	.246	.536	.846

Table 26

Median *P*-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Highest Single Level, Grade 8

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	2	.601	.774	.895	.953
Basic	13	.409	.842	.919	.948
Proficient	121	.280	.529	.828	.949
Advanced	75	.371	.649	.844	.942

Table 27

Median *P*-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Highest Single Level, Grade 12

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	34	.315	.613	.848	.959
Basic	88	.353	.667	.879	.968
Proficient	73	.296	.554	.807	.940
Advanced	13	.091	.142	.552	.855

In many cases, students who scored in either the Basic or Proficient level performed equally well or better on the set of items assigned to the next highest level. For example, Figure 5 shows that at Grade 4, students scoring in the Basic range perform as well on Proficient items (median *p*-value = .475) as they do on Basic items (median *p*-value = .471). Figure 6 shows that 8th-grade students classified at the Proficient level performed better on the set of Advanced items (median *p*-value = .844) than they did on Proficient items (median *p*-value = .828).

Assuming that the sets of items mapped to descriptors really do call for the knowledge and skills referred to in the descriptors, then one is forced to conclude that the descriptions do not provide a clear indication of which items students at a given level are likely to be able to answer correctly.

Conclusions and Recommendations

The goal of making meaningful interpretations of performance on the NAEP has led to a number of attempts to interpret scores on the NAEP mathematics score scale in terms of what students with particular scores can or should be able to do (Phillips et al., 1993). Any approach adopted towards this goal will be open to criticism unless the criteria by which performance is judged are well-defined. According to Nitko (1984), "a domain is well defined if it is clear which categories of performance or which kinds of tasks are and are not potential test items" (p. 12). The lack of agreement among the mathematics educators in this study on the items that would measure many of the skills

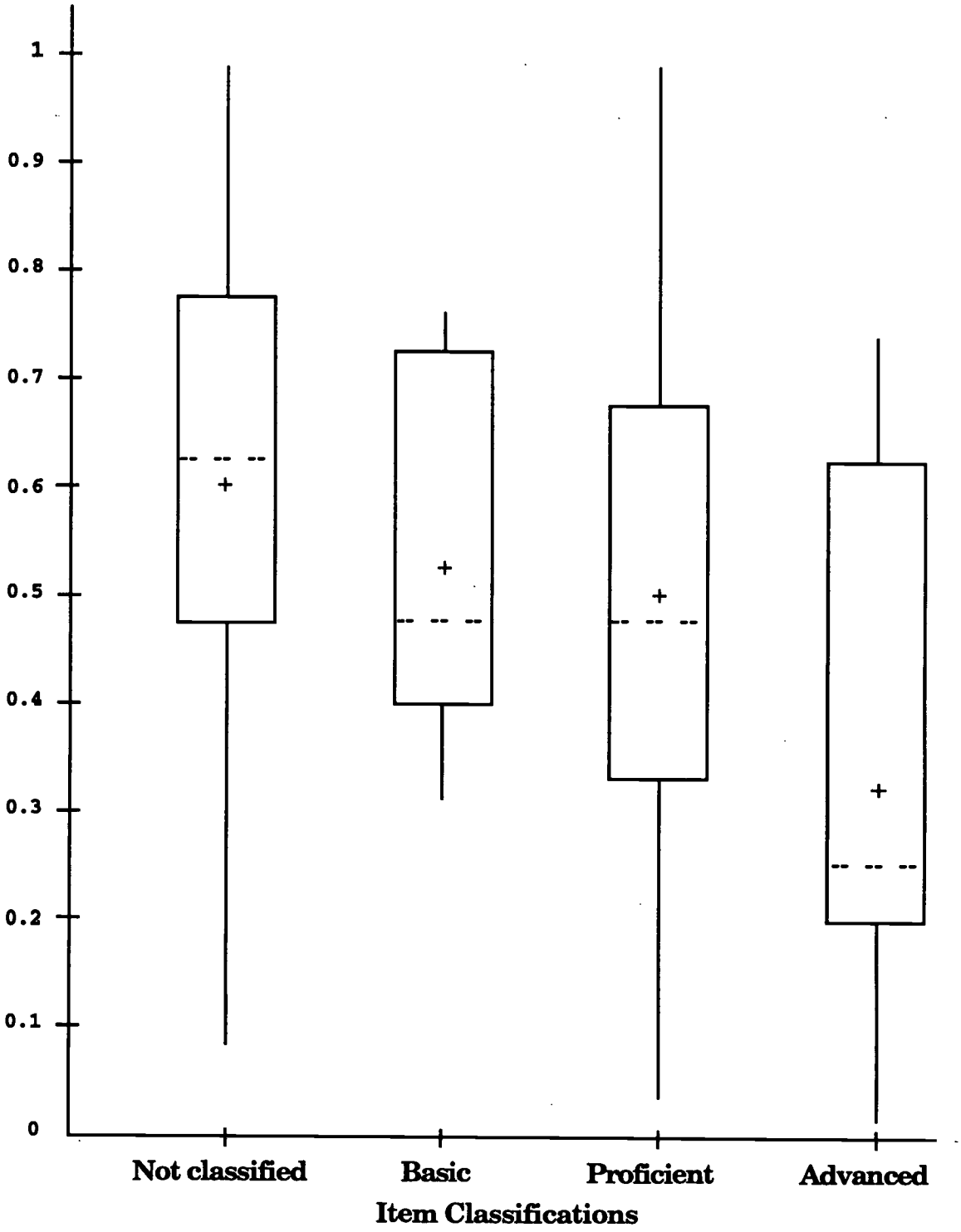


Figure 5. Distribution of item percents correct (*p*-values) at each level, for Basic students, on sets of items not classified, or classified as Basic, Proficient, and Advanced based on judges' mappings, Grade 4.

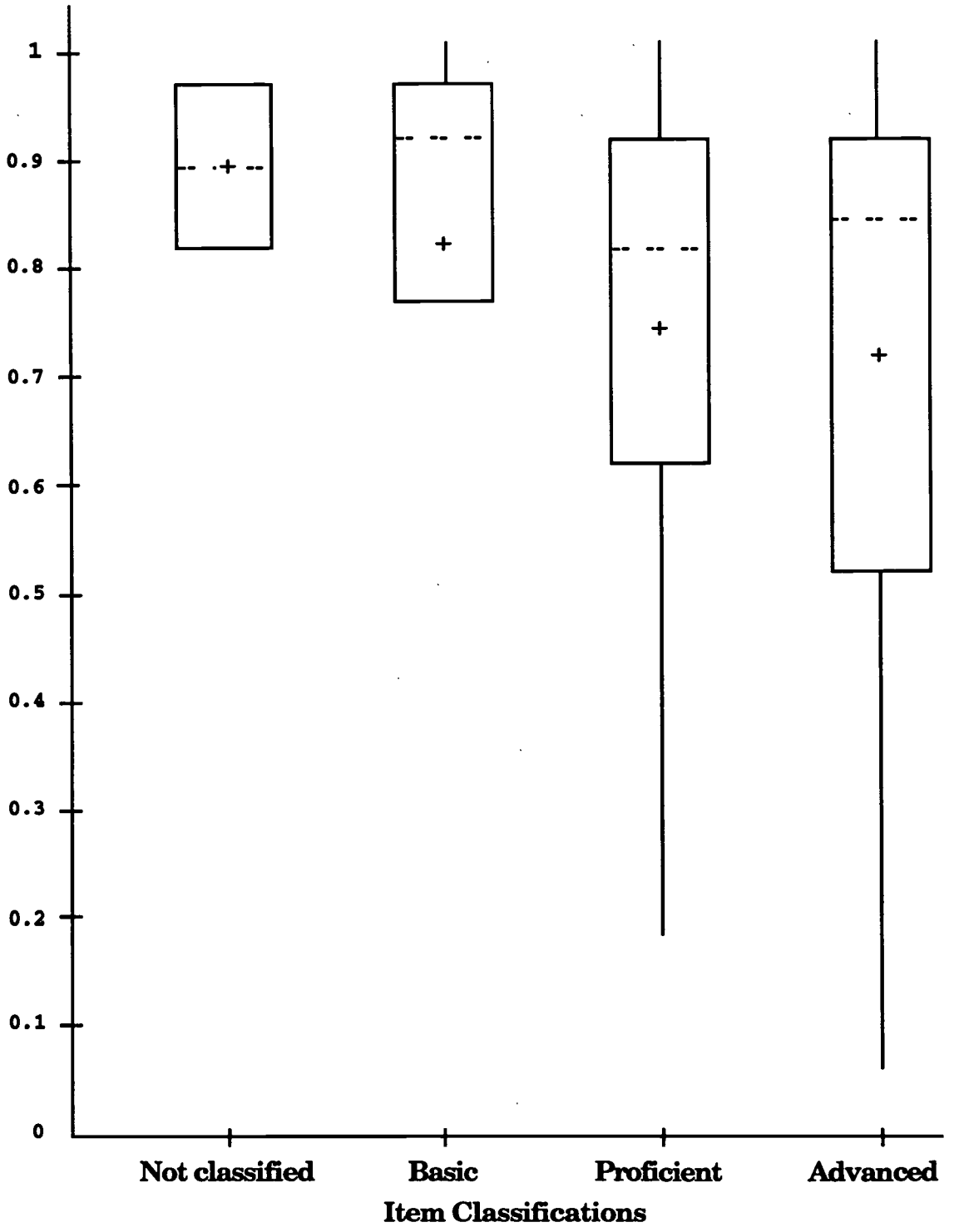


Figure 6. Distribution of item percents correct (*s*-values) at each level, for Proficient students, on sets of items not classified, or classified as Basic, Proficient, and Advanced based on judges' mappings, Grade 8.

included in the narrative descriptions indicates that the criteria to which NAEP is attempting to relate performance are too ill-defined. Forsyth (1991) made a similar claim about the NAEP scales used to interpret performance on the 1986 NAEP in mathematics.

For sets of items that were consistently judged to tap some of the performances included in the 1992 achievement level descriptions, the performance of students on those sets of items was not consistent with the levels to which they (the students) were assigned based on their NAEP scores. Therefore, it is not appropriate to conclude that students scoring at particular levels know and are able to do what NAGB's achievement level descriptions indicate they should be able to do.

In a criterion-referenced assessment system, both the development of items and the interpretation of performance should be driven by the assessment objectives framework, the objectives serving as the criteria for judging performance (Nitko, 1984). The objectives framework used to generate the 1992 NAEP mathematics item pool had two dimensions: content and mathematical ability. The NAGB descriptions mix these two dimensions of performance. NAEP has never reported scores across items targeting particular mathematical abilities, either within content areas or independent of content areas. By including references to the ability dimensions in the achievement level descriptions, NAEP is perhaps committing itself to analysis and reporting on aspects of performance that may be difficult, if not impossible, to isolate or tie to particular items or sets of items. Our analysis of the achievement level descriptions was premised on the notion that it is indeed possible to identify both the content and mathematical ability features of the test items.

One important step for NAGB to adopt in establishing achievement levels in mathematics would be to start the process anew by closely aligning the characterization of achievement levels with the development of the new assessment frameworks, items, and associated data collection. Linking level setting with assessment design from the outset may provide the only means to determine whether it is possible to develop valid descriptions of what students know and can do.

References

Bourque, M.L. (1993, April). *The NAEP achievement level setting process for the 1992 mathematics assessment*. Paper presented at a joint symposium of the American Educational Research Association and the National Council for Measurement in Education annual meeting, Atlanta, GA.

Bourque, M.L., & Garrison, H.H. (1991). *The levels of mathematics achievement. Vol. I. National and state summaries.* Washington, DC: National Assessment Governing Board.

Burstein, L., Koretz, D.M., Linn, R L., Sugrue, B., Novak, J., Lewis, E., & Baker, E.L. (1993). *The validity of interpretations of the 1992 NAEP achievement levels in mathematics* (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Educational Testing Service/National Assessment of Educational Progress. (1988). *Mathematics objectives, 1990 assessment*. Princeton, NJ: Author.

Forsyth, R.A. (1991). Do NAEP scales yield valid criterion-referenced interpretations? *Educational Measurement: Issues and Practice, 10*, 3-16.

French, J.W. (1965). The relationship of problem-solving styles to the factor composition of tests. *Educational and Psychological Measurement, 25*, 9-28.

General Accounting Office. (1993). *Educational achievement standards: NAGB's approach yields misleading interpretations.* Washington, DC: Author.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21. (ERIC Document Reproduction Service No. EJ 436 999)

Linn, R.L., Koretz, D.M., Baker, E.L., & Burstein, L. (1991). *The validity and credibility of the achievement levels for the 1990 National Assessment of Educational Progress in mathematics* (CSE Tech. Rep. No. 330). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Mullis, I.V.S., Dossey, J.A., Owen, E.H., & Phillips, G.W. (1993). *NAEP 1992 mathematics report card for the nation and the states*. Washington, DC: U.S. Department of Education.

National Academy of Education. (1993). *Setting performance standards for student achievement. A report of the National Academy of Education*

panel of the evaluation of the NAEP trial state assessment: An evaluation of the 1992 achievement levels. Stanford, CA: NAE, Stanford University.

National Assessment Governing Board. (1991). *Response to the draft summative evaluation report on the National Governing Board's inaugural effort to set achievement levels on the National Assessment of Educational Progress.* Washington, DC: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Nitko, A.J. (1984). Defining "criterion-referenced test". In R.A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 8-28). Baltimore, MD: Johns Hopkins University Press.

Novak, J., Burstein, L., & Larriva, C. (forthcoming). *Characteristics of items that differentiate among students classified at different levels of achievement in the 1992 NAEP in mathematics* (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Novak, J., Burstein, L., & Sugrue, B. (forthcoming). *Sources of variability in mathematics educators' mapping of achievement level descriptions to 1992 NAEP mathematics test items* (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Phillips, G.W., Mullis, I.V.S., Bourque, M.L., Williams, P.L., Hambleton, R.K., Owen, E.H., & Barton, P.E. (1993). *Interpreting NAEP scales.* Washington, DC: U.S. Department of Education.

Reckase, M.D. (1993). *The defensibility of domain descriptions for achievement levels and anchor points.* Paper developed for the meeting of the Technical Advisory Committee on Standard Setting (TACSS) in support of the methodology devised by ACT for the National Assessment Governing Board, Washington, DC, February 9-10, 1993.

Stufflebeam, D.L., Jaeger, R.M., & Scriven, M. (1991). *Summative evaluation of the National Governing Board's inaugural effort to set achievement levels on the National Assessment of Educational Progress.* Kalamazoo: Western Michigan University.

Appendix A

Original Narrative Descriptions of the NAEP's 1992 Mathematics Achievement Levels for Grades 4, 8, and 12*

NAEP Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Fourth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra and functions. At the fourth-grade level, algebra and functions are treated in informal and exploratory ways, often through the study of patterns. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 211	Fourth-grade students performing at the basic level should show some evidence of understanding the mathematical concepts and procedures in the five NAEP content areas.

Fourth graders performing at the level should be able to estimate and use basic facts to perform simple computations with whole numbers; show some understanding of fractions and decimals; and solve some simple real-world problems in all NAEP content areas. Students at this level should be able to use—though not always accurately—four-function calculators, rulers, and geometric shapes. Their written responses are often minimal and presented without supporting information.

* SOURCE: Figure 1.3, Mullis, I.V.S. et al., (1993), *NAEP 1992 Mathematics Report Card for the Nation and the States*, pp. 44, 51, and 56.

Proficient 248 Fourth-grade students performing at the proficient level should consistently apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas.

Fourth graders performing at the proficient level should be able to use whole numbers to estimate, compute, and determine whether results are reasonable. They should have a conceptual understanding of fractions and decimals; be able to solve real-world problems in all NAEP content areas; and use four-function calculators, rulers, and geometric shapes appropriately. Students performing at the proficient level should employ problem-solving strategies such as identifying and using appropriate information. Their written solutions should be organized and presented both with supporting information and explanations of how they were achieved.

Advanced 280 Fourth-grade students performing at the advanced level should apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas.

Fourth graders performing at the advanced level should be able to solve complex and nonroutine real-world problems in all NAEP content areas. They should display mastery in the use of four-function calculators, rulers, and geometric shapes. These students are expected to draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved. They should go beyond the obvious in their interpretations and be able to communicate their thoughts clearly and concisely.

NAEP Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Eighth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 256 — Eighth-grade students performing at the basic level should exhibit evidence of conceptual and procedural understanding in the five NAEP content areas. This level of performance signifies understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.

Eighth graders performing at the basic level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools—including calculators, computers, and geometric shapes. Students at this level should also be able to use fundamental algebraic and informal geometric concepts in problem solving.

As they approach the proficient level, students at the basic level should be able to determine which of available data are necessary and sufficient for correct solutions and use them in problem solving. However, these 8th graders show limited skill in communicating mathematically.

Proficient 294 Eighth-grade students performing at the proficient level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content areas.

Eighth graders performing at the proficient level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical solutions.

Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs; apply properties of informal geometry; and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.

Advanced 331 Eighth-grade students performing at the advanced level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content areas.

Eighth graders performing at the advanced level should be able to probe examples and counter-examples in order to shape generalizations from which they can develop models. Eighth graders performing at the advanced level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions.

Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Twelfth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 287	Twelfth-grade students performing at the basic level should demonstrate procedural and conceptual knowledge in solving problems in the five NAEP content areas.

Twelfth-grade students performing at the basic level should be able to use estimation to verify solutions and determine the reasonableness of results as applied to real-world problems. They are expected to use algebraic and geometric reasoning strategies to solve problems. Twelfth graders performing at the basic level should recognize relationships presented in verbal, algebraic, tabular, and graphical forms; and demonstrate knowledge of geometric relationships and corresponding measurement skills.

They should be able to apply statistical reasoning in the organizations and display of data and in reading tables and graphs. They should be able to generalize from patterns and examples in the areas of algebra, geometry, and statistics. At this level, they should use correct mathematical language and symbols to communicate mathematical relationships and reasoning processes; and use calculators appropriately to solve problems.

Proficient 334 Twelfth-grade students performing at the proficient level should consistently integrate mathematical concepts and procedures to the solutions of more complex problems in the five NAEP content areas.

Twelfth-grade students performing at the proficient level should demonstrate an understanding of algebraic, statistical, and geometric and spatial reasoning. They should be able to perform algebraic operations involving polynomials; justify geometric relationships; and judge and defend the reasonableness of answers as applied to real-world situations. These students should be able to analyze and interpret data in tabular and graphic form; understand and use elements of the function concept in symbolic, graphical, and tabular form; and make conjectures, defend ideas, and give supporting examples.

Advanced 366 Twelfth-grade students performing at the advanced level should consistently demonstrate the integration of procedural and conceptual knowledge and the synthesis of ideas in the five NAEP content areas.

Twelfth-grade students performing at the advanced level should understand the function concept; and be able to compare and apply the numeric, algebraic, and graphical properties of functions. They should apply their knowledge of algebra, geometry, and statistics to solve problems in more advanced areas of continuous and discrete mathematics.

They should be able to formulate generalizations and create models through probing examples and counter examples. They should be able to communicate their mathematical reasoning through the clear, concise, and correct use of mathematical symbolism and logical thinking.

Appendix B

Parsed Versions of the NAEP Achievement Level Descriptions

NAEP Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Fourth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra and functions. At the fourth-grade level, algebra and functions are treated in informal and exploratory ways, often through the study of patterns. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 211

1. Fourth-grade students performing at the basic level should show *some evidence of understanding the mathematical concepts and procedures* in the five NAEP content areas.

2. Fourth graders performing at the level should be able to
   a. *estimate with whole numbers.*
   b. *use basic facts to perform simple computations with whole numbers*;
3. show
   a. *some understanding of fractions*
   b. *some understanding of decimals*;

4. and *solve some simple real-world problems* in all NAEP content areas.

5. Students at this level should be able to *use—though not always accurately—four-function calculators, rulers, and geometric shapes.*

6. Their *written responses* are often
   a. *minimal*
   b. and presented *without supporting information.*

Proficient 248

7. Fourth-grade students performing at the proficient level should *consistently apply integrated procedural knowledge and conceptual understanding to problem solving* in the five NAEP content areas.

8. Fourth graders performing at the proficient level should be able *to use whole numbers to*

a. *estimate results,*

b. *compute results,*

c. *determine whether results are reasonable.*

9. They should have a

a. *conceptual understanding of fractions*

b. *conceptual understanding of decimals*;

10. be able to *solve real-world problems* in all NAEP content areas;

11. *use four-function calculators, rulers, and geometric shapes appropriately.*

12. *should employ problem-solving strategies such as identifying and using appropriate information.*

13. Their *written solutions* should be

a. *organized*

b. *presented* both *with supporting information*

c. *presented with explanations of how they were achieved.*

Advanced 280

14. Fourth-grade students performing at the advanced level should *apply integrated procedural knowledge and conceptual understanding to complex and non-routine real-world problem solving* in the five NAEP content areas.

15. Fourth graders performing at the advanced level should be able to *solve complex and nonroutine real-world problems* in all NAEP content areas.

16. They should display *mastery in the use of four-function calculators, rulers, and geometric shapes.*

17. These students are expected to *draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved.*

18. They should

a. *go beyond the obvious in their interpretations*

b. and be able to *communicate their thoughts clearly*

c. and *communicate their thoughts concisely.*

NAEP Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Eighth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 256

1. Eighth-grade students performing at the basic level should exhibit *evidence of conceptual and procedural understanding* in the five NAEP content areas.

2. This level of performance signifies *understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.*

3. Eighth graders performing at the basic level should *complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs.*

4. They should be able to *solve problems* in all NAEP content areas
   a. *through the appropriate selection and use of strategies*
   b. *the appropriate selection and use and technological tools including calculators, computers, and geometric shapes.*

5. Students at this level should also be able to
   a. use fundamental algebraic concepts in problem solving.
   b. and use informal geometric concepts in problem solving.

6. As they approach the proficient level, students at the basic level should be able to
   a. *determine which of available data are necessary and sufficient for correct solutions*
   b. and *use them [data] in problem solving.*

7. However, these 8th graders show *limited skill in communicating mathematically.*

Proficient 294

8. **Eighth-grade students performing at the proficient level should *apply mathematical concepts and procedures consistently to complex problems* in the five NAEP content areas.**

9. Eighth graders performing at the proficient level should be able to
   a. *conjecture,*
   b. *defend their ideas,*
   c. and *give supporting examples.*

10. They should
   a. *understand the connections between fractions, percents, decimals,*
   b. and *[connections between] other mathematical topics such as algebra and functions.*

11. Students at this level are expected to have a *thorough understanding of basic-level arithmetic operations—an understanding sufficient for problem solving in practical solutions.*

12. *Quantity and spatial relationships in problem solving and reasoning* should be familiar to them,

13. and they should be able to *convey underlying reasoning skills beyond the level of arithmetic.*

14. They should be able to
   a. *compare and contrast mathematical ideas* and
   b. *generate their own examples.*

15. These students should *make inferences from data and graphs;*

16. *apply properties of informal geometry;*

17. and *accurately use the tools of technology.*

18. Students at this level should
   a. *understand the process of gathering and organizing data*
   b. and be able to *calculate and evaluate results within the domain of statistics and probability.*
   c. and *communicate results within the domain of statistics and probability.*

Advanced 331

19. **Eighth-grade students performing at the advanced level should be able to**

   a. ***reach beyond the recognition, identification, and application of mathematical rules in order to generalize***

   b. **and *synthesize concepts and principles* in the five NAEP content areas.**

20. Eighth graders performing at the advanced level should be able to *probe examples and counter-examples in order to shape generalizations from which they can develop models.*

21. Eighth graders performing at the advanced level should

   a. *use number sense to consider the reasonableness of an answer.*

   b. and *use geometric awareness to consider the reasonableness of an answer.*

22. They are expected to

   a. *use abstract thinking to create unique problem-solving techniques*

   b. and *explain the reasoning processes underlying their conclusions.*

Description of Mathematics Achievement Levels for Basic, Advanced, and Proficient Twelfth Graders

The five NAEP content areas are (1) numbers and operations, (2) measurement, (3) geometry, (4) data analysis, statistics, and probability, and (5) algebra functions. Skills are cumulative across levels—from Basic to Proficient to Advanced.

Basic 287

1. Twelfth-grade students performing at the basic level should *demonstrate procedural and conceptual knowledge in solving problems* in the five NAEP content areas.

2. Twelfth-grade students performing at the basic level should be able to *use estimation* to

   a. *verify solutions as applied to real-world problems*

   b. and *determine the reasonableness of results as applied to real-world problems.*

3. They are expected to

   a. *use algebraic reasoning strategies to solve problems.*

   b. and *use geometric reasoning strategies to solve problems.*

4. Twelfth graders performing at the basic level should *recognize relationships presented in verbal, algebraic, tabular, and graphical forms*;

5. and *demonstrate knowledge of geometric relationships and corresponding measurement skills.*

6. They should be able to *apply statistical reasoning*

   a. *in the organization and display of data*

   b. and *in reading tables and graphs.*

7. They should be able to

   a. *generalize from patterns and examples in the area of algebra,*

   b. *generalize from patterns and examples in the area of geometry,*

   c. *generalize from patterns and examples in the area of statistics.*

8. At this level, they should

   a. *use correct mathematical language and symbols to communicate mathematical relationships*

   b. and *use correct mathematical language and symbols to communicate mathematical reasoning processes*;

9. *use calculators appropriately to solve problems.*

Proficient 334

10. Twelfth-grade students performing at the proficient level should *consistently integrate mathematical concepts and procedures to the solutions of more complex problems* in the five NAEP content areas.

11. Twelfth-grade students performing at the proficient level should

   a. *demonstrate an understanding of algebraic reasoning.*

   b. *demonstrate an understanding of statistical reasoning.*

   c. *demonstrate an understanding of geometric and spatial reasoning.*

12. They should be able to *perform algebraic operations involving polynomials;*

13. *justify geometric relationships*;

14. and *judge and defend the reasonableness of answers as applied to real-world situations.*

15. These students should be able to *analyze and interpret data in tabular and graphical form;*

16. *understand the elements of the function concept in symbolic, graphical, and tabular form;*

17. and *use elements of the function concept in symbolic, graphical, and tabular form;*

18. and

   a. *make conjectures,*

   b. *defend ideas,*

   c. and *give supporting examples.*

19. **Twelfth-grade students performing at the advanced level should**
    a. ***consistently demonstrate the integration of procedural and conceptual knowledge***
    b. **and *consistently demonstrate the synthesis of ideas* in the five NAEP content areas.**

20. Twelfth-grade students performing at the advanced level should *understand the function concept*;

21. and be able to
    a. *compare the numeric, algebraic, and graphical properties of functions.*
    b. and *apply the numeric, algebraic, and graphical properties of functions.*

22. They should a*pply their knowledge of algebra, geometry, and statistics to solve problems in more advanced areas of continuous and discrete mathematics.*

23. They should be able to *formulate generalizations and create models through probing examples and counter examples.*

24. They should be able to *communicate their mathematical reasoning* through the *clear, concise, and correct use of mathematical symbolism and logical thinking*,

Appendix C

Final Versions of Descriptors Used to Map NAEP Assessment Items to Levels

Grade 4 Descriptors

Block _____ Item _____ Item ID _____

Match each test item to as many of the following descriptions as appropriate. If a description applies to an item, put a check mark in the LINE to the left of the description. Also, if you are NOT sure of any decision (whether checked or left blank), circle the "?" to the right of the description.

1. ___ If the item involves whole numbers, check any of the following descriptions that apply:

   The item calls for:

   1(a) ____ using basic number facts to perform simple computations with whole numbers ?

   1(b) ____ estimating with whole numbers ?

   1(c) ____ using whole numbers to compute results ?

   1(d) ____ using whole numbers to estimate results ?

   1(e) ____ determining of the reasonableness of whole number results ?

2. ___ If the item involves fractions or decimals, indicate which one of the following descriptions best applies to the item:

   The item calls for:

   2(a) ____ some understanding of fractions or decimals ?

   or

   2(b) ____ conceptual understanding of fractions or decimals ?

3. ___ The item calls for understanding of mathematical concepts or mathematical procedures. ?

4. ___ The item calls for applying integrated procedural and conceptual understanding to problem solving ?

5. ___ The item calls for applying integrated procedural and conceptual understanding to complex and nonroutine real-world problem solving ?

6. ___ If the item calls for real-world problem-solving, check which one of the following best describes the item:

The item calls for:

6(a) ____ solving a simple real-world problem ?

or

6(b) ____ solving a [routine] real-world problem ?

or

6(c) ____ solving a complex and nonroutine real-world problem ?

7. ___ The item calls for employing problem-solving strategies such as identifying and using appropriate information ?

8. ___ If the item calls for a written response, check any of the following descriptions that apply:

The item calls for:

8(a) ____ giving supporting information ?

8(b) ____ explaining how the answer or solution process was achieved ?

8(c) ____ explaining why the answer or solution process was achieved ?

8(d) ____ clear or concise communication ?

Grade 8 Descriptors

Block _____ Item _____ Item ID _____

Match each test item to as many of the following descriptions as appropriate. If a description applies to an item, put a check mark in the LINE to the left of the description. Also, if you are NOT sure of any decision (whether checked or left blank), circle the "?" to the right of the description.

1. ___ The item calls for an understanding of arithmetic operations—including estimation — on whole numbers, decimals, fractions or percents. ?

2. ___ The item calls for a thorough understanding of basic-level arithmetic operations — an understanding sufficient for problem solving in practical situations. ?

3. ___ The item calls for understanding the connections among any of the following: fractions, percents, decimals. ?

4. ___ The item calls for using fundamental algebraic concepts in problem solving. ?

5. ___ The item calls for understanding of the connection between algebra and functions. ?

6. ___ If the item involves geometric concepts, check any of the following descriptions that apply:

   The item calls for:

   6(a) ____ using informal geometric concepts in problem solving ?

   6(b) ____ applying the properties of informal geometry ?

   6(c) ____ using geometric awareness to consider the reasonableness of an answer ?

7. ___ The item calls for familiarity with quantity or spatial relationships in problem solving or reasoning. ?

8. ___ The item calls for completing problems with the help of structural prompts such as diagrams, charts, or graphs. ?

9. ___ The item calls for solving problems through the appropriate selection and use of strategies. ?

10. ___ The item calls for solving problems through the appropriate selection and use of technological tools—including calculators, computers, or geometric shapes. ?

11. ___ The item calls for using abstract thinking to create unique problem-solving techniques. ?

12. ___ The item calls for determining which of available data are necessary and sufficient for correct solutions. ?

13. ___ If the item involves working with data, check any of the following descriptions that apply:

The item calls for:

13(a) ___ making of inferences from data or graphs ?

13(b) ___ understanding of the process of gathering and organizing data ?

14. ___ If the item involves statistics or probability, check any of the following descriptions that apply:

The item calls for:

14(a) ___ calculating results within the domain of statistics or probability ?

14(b) ___ evaluating results within the domain of statistics or probability ?

14(c) ___ communicating results within the domain of statistics or probability ?

15. ___ The item calls for conceptual understanding or procedural understanding ?

16. ___ The item calls for applying mathematical concepts and procedures to complex problems. ?

17. ___ The item calls for reaching beyond the recognition, identification, and
application of mathematical rules to generalize and synthesize concepts and principles. ?

18. ___ The item calls for comparing and contrasting mathematical ideas. ?

19. ___ The item calls for generating one's own examples. ?

20. ___ The item calls for probing of examples and counter examples in order to shape generalizations from which the student can develop models. ?

21. ___ The item calls for the using number sense to consider the reasonableness of an answer. ?

22. ___ If the item requires a written response, check any of the following descriptions that apply:

The item calls for:

22(a) ____ making conjectures ?

22(b) ____ defending ideas ?

22(c) ____ giving supporting examples ?

22(d) ____ explaining the reasoning process underlying conclusions ?

22(e) ____ conveying underlying reasoning skills beyond the level of arithmetic ?

Grade 12 Descriptors

Block _____ Item _____ Item ID _____

Match each test item to as many of the following descriptions as appropriate. If a description applies to an item, put a check mark in the LINE to the left of the description. Also, if you are NOT sure of any decision (whether checked or left blank), circle the "?" to the right of the description.

1. ___ If the item involves geometry, check any of the following descriptions that apply:

   The item calls for:

   1(a) ____ using geometric reasoning strategies to solve problems ?

   1(b) ____ knowledge of geometric relationships and corresponding measurement skills ?

   1(c) ____ an understanding of geometric reasoning ?

   1(d) ____ an understanding of spatial reasoning ?

   1(e) ____ justifying geometric relationships ?

   1(f) ____ generalizing from patterns or examples ?

2. ___ If the item involves algebra, check any of the following descriptions that apply:

   The item calls for:

   2(a) ____ using algebraic reasoning strategies to solve problems ?

   2(b) ____ an understanding of algebraic reasoning ?

   2(c) ____ performing algebraic operations involving polynomials ?

   2(d) ____ generalizing from patterns or examples ?

3. ___ If the item involves functions, check any of the following descriptions that apply:

   The item calls for:

   3(a) ____ understanding of elements of the function concept in symbolic, graphical or tabular form ?

3(b) ____ understanding of the function concept ?

3(c) ____ using elements of the function concept in symbolic, graphical or tabular form ?

3(d) ____ comparing the numeric, algebraic, or graphical properties of functions ?

3(e) ____ applying the numeric, algebraic, or graphical properties of functions ?

4. ___ If the item involves data analysis or statistics, check any of the following descriptions that apply:

The item calls for:

4(a) ____ applying statistical reasoning in the organization and display of data ?

4(b) ____ applying statistical reasoning in reading tables or graphs ?

4(c) ____ an understanding of statistical reasoning ?

4(d) ____ analyzing and interpreting data in tabular or graphical form ?

4(e) ____ generalizing from patterns or examples ?

5. ___ The item calls for solution of problems in the more advanced area of continuous and discrete mathematics. ?

6. ___ The item calls for recognizing relationships presented in verbal, algebraic, tabular, or graphical forms. ?

7. ___ The item calls for formulating generalizations and creating models through probing examples and counterexamples. ?

8. ___ The item calls for using estimation to verify solutions to real-world problems. ?

9. ___ The item calls for using estimation to determine the reasonableness of results as applied to real-world problems. ?

10. ___ The item calls for judging or defending the reasonableness of answers as applied to real-world situations. ?

11. ___ The item calls for procedural knowledge or conceptual knowledge in solving problems. ?

12. ___ The item calls for integrating mathematical concepts and procedures to the solution of more complex problems. ?

13. ___ The item calls for the integration of procedural and conceptual knowledge, and the synthesis of ideas. ?

14. ___ If the item requires a written response, check any of the following descriptions that apply ?

The item calls for:

14(a) ____ using mathematical language and symbols to communicate mathematical relationships ?

14(b) ____ using mathematical language and symbols to communicate reasoning processes. ?

14(c) ____ clear and concise use of mathematical symbolism and logical thinking to communicate mathematical reasoning. ?

14(d) ____ defending ideas ?

14(e) ____ making conjectures ?

14(f) ____ giving supporting examples ?

Appendix D

Judges' Background Questionnaire and Summary of Data

The information you provide on this form will be held strictly confidential. It is important that we have detailed descriptions of the characteristics of the group of mathematics educators who act as judges on this project. We will not be reporting or making judgments about any individual participant. Thank you for your cooperation.

1. **Name** _____

2. **Current Position (Title, location, description of responsibilities):**

   _____

3. **How long have you been at the current position?** _____

4. **What is your sex? (Circle one.)**

   Male 1
   Female 2

5. **Which best describes you? (Circle one.)**

   American Indian or Alaskan Native 1
   Asian or Pacific Islander 2
   Hispanic, regardless of race 3
   Black (not of Hispanic origin) 4
   White (not of Hispanic origin) 5
   Other (specify) _____ 6

6. **Please list the degrees you hold, your major field of study, and the institution and year you obtained each degree:**

   If you have not completed a degree, check here (____) and go to Question 7.

	Degree	Major	Institution	Year
Bachelor's Degree	_____	_____	_____	_____
Master's Degree	_____	_____	_____	_____
Doctorate (e.g., Ed.D., Ph.D.)	_____	_____	_____	_____

7. **Please indicate the level of exposure you have had, if any, to each of the following topics or areas:**

Topic/Area	One or more college or university courses	Part of a college or university course	In-service training	Little or no exposure
Methods of teaching mathematics	_____	_____	_____	_____
Number systems and numeration	_____	_____	_____	_____
Measurement in mathematics	_____	_____	_____	_____
Geometry	_____	_____	_____	_____
Probability/statistics	_____	_____	_____	_____
Abstract/linear algebra	_____	_____	_____	_____
Calculus	_____	_____	_____	_____
Psychology of learning	_____	_____	_____	_____
Cognitive psychology	_____	_____	_____	_____

8. **Have you ever had training in any of the following, either in college courses or in in-service education? (Circle Yes or No for each area).**

Estimation	YES	NO
Problem-solving in mathematics	YES	NO
Use of manipulatives (e.g., measuring instruments or geometric solids) in mathematics education	YES	NO
Use of calculators in mathematics instruction	YES	NO
Understanding students' thinking about mathematics	YES	NO
Gender issues in the teaching of mathematics	YES	NO
Teaching students from different cultural backgrounds	YES	NO

9. **Please indicate the number of years you have taught mathematics at each of the following school levels:**

School Level	Number of Years
Prekindergarten	_____
Elementary (K-5)	_____
Middle/Junior High (6-8)	_____
Senior High (9-12)	_____
Postsecondary	_____

10. **In which subject areas do you hold a current state teaching certificate? (Circle ALL that apply.)**

Elementary education (specify grades) _____	01
Middle school education (specify grades) _____	02
High school education (specify grades) _____	03
General science	04
Biology, environmental/life sciences	05
Earth/space sciences	06
Physical sciences	07
Chemistry	08
Physics	09
Mathematics	10
Computer science	11
Business	12
English/language arts, reading	13
Physical education, health	14
Social studies	15
Other (specify) _____	16

11. **During the last 3 years, what is the total amount of time you have spent on in-service education in mathematics or the teaching of mathematics? (Include attendance at professional meetings, workshops, and conferences, but do not include formal courses for which you received college credit.) (Circle one.)**

None	1
Less than 6 hours	2
6 to 15 hours	3
16 to 35 hours	4
More than 35 hours	5

12. **How familiar are you with the emphasis of the Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989)? (Circle one.)**

A. Very familiar

B. Somewhat familiar

C. Not familiar

13. **How familiar are you with the emphasis of the *California Mathematics Framework* (CDE, 1985, 1992)? (Circle one.)**

A. Very familiar

B. Somewhat familiar

C. Not familiar

14. **During the past two years, how often did you present or participate in CMC or National Council of Teachers of Mathematics (NCTM) conferences, Replacement Unit workshops, or district curriculum improvement workshops? (Circle one.)**

A. Did not participate.

B. Participated once or twice

C. Participated three or four times

D. Participated more than four times.

15. **Did you participate in a California Mathematics Project or projects such as EQUALS? (Circle one.)**

A. Yes

B. No

16. **If you are currently teaching mathematics at <u>middle or high</u> school level, indicate, for each mathematics course you teach, the title of the course, the course level, the ability level of students in that class, and the number of years you have taught this type of class:**

	Course 1	2	3	4
Course Title	_____	_____	_____	_____
	_____	_____	_____	_____
Grade Level	_____	_____	_____	_____
Course Level Select one: 1. Enriched/AP 2. General/Regular 3. Remedial 4. Applied/vocational	_____	_____	_____	_____
Student math ability level Select one: 1. Primarily high 2. Primarily average 3. Primarily low 4. Widely mixed	_____	_____	_____	_____
Number of years you have taught this type of class	_____	_____	_____	_____

SUMMARY OF CHARACTERISTICS OF THE JUDGES

A background and teaching experience questionnaire given to the 18 raters revealed the following information:

Personal History: There were 13 females and five males.
Ethnic representation included 9 Caucasians, 5 African-Americans, 3 Hispanics, and 1 Asian.
All but one were currently in a teaching position at the time of the judging (the one exception was working as a clinical consultant for secondary mathematics in the UCLA teacher training program.)

Education Level: Every judge held a bachelors degree, eleven of which were in the fields of math, science or engineering, three in education, and six in other fields. Four held masters degrees and two had doctorates.

Years Teaching Math: Judges' mathematics teaching experience ranged from 1 to 33 years experience at the Elementary level, 1 to 9 years at Middle/Jr. High level, and 1 to 16 years at the Sr. High level. The mean number of years of mathematics teaching experience was 12.2 (median 12).

Certification: Every judge held a current teaching credential. Ten of the eighteen held credentials in mathematics, six in high school education, 8 in middle school education and 9 in elementary education.

Exposure to Topics through University Courses or In-service: All judges (100%) had exposure to the following topic areas: methods of teaching math, numeration, measurement, problem solving, manipulatives, psychology of learning and teaching students from various cultures. All but one had exposure in geometry and probability.

The majority (at least 83%) had received training in the use of calculators, in the understanding of students' thinking about mathematics and in estimation. The majority had also spent more than 35 hours during the last 3 years on in-service education in the teaching of math.

Familiarity with Mathematics Standards: The majority (83%) were familiar with NCTM Curriculum and Evaluation Standards for School Mathematics and every judge was familiar with the California Mathematics Framework.

Conference Workshop Participation: Seventeen of the eighteen judges participated or presented at national or district conferences and/or workshops during the past two years.

Appendix E

Computer Screens for Grade 8 Mapping Protocol

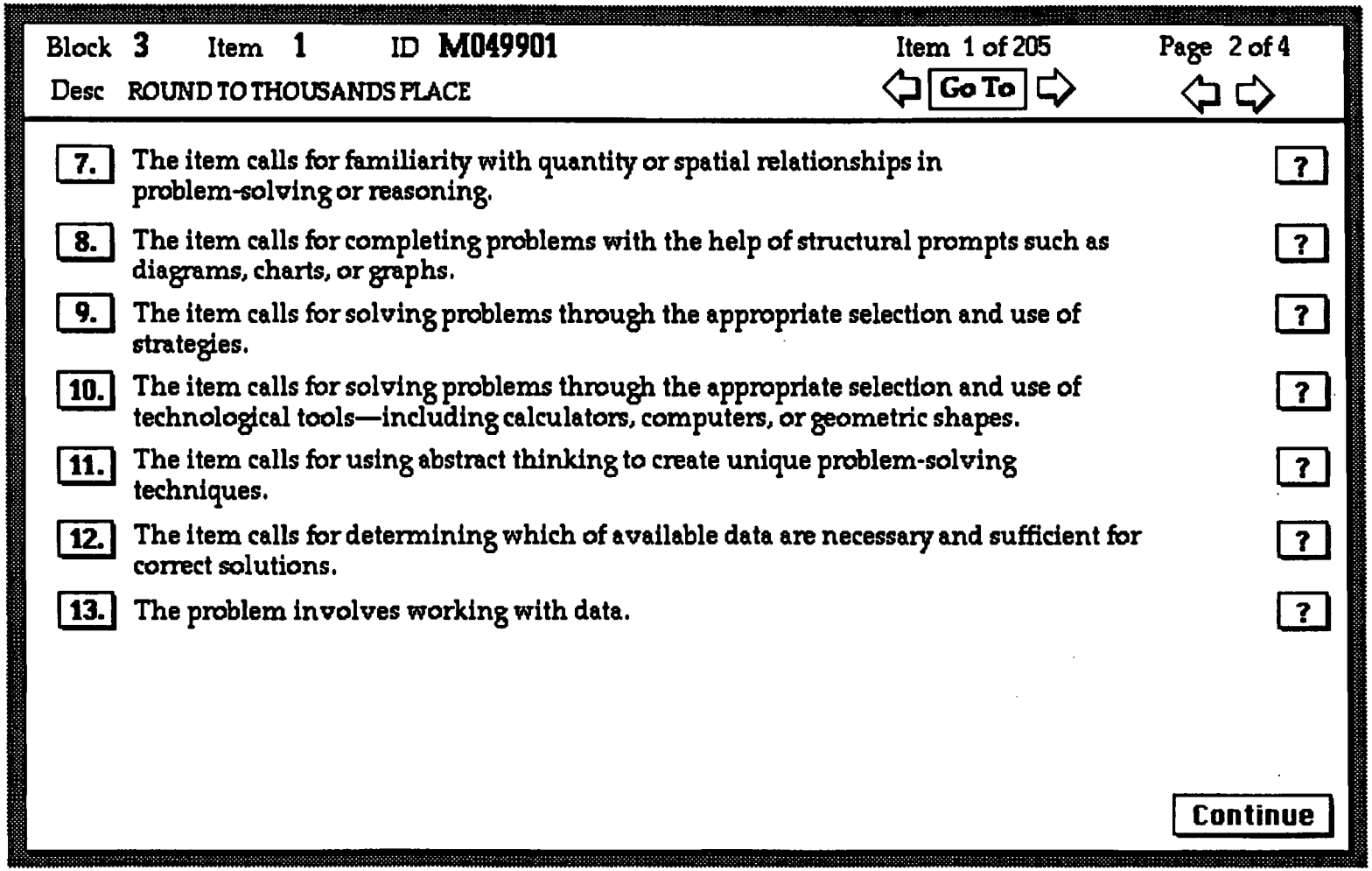
Block 3 Item 1 ID M049901 Item 1 of 205 Page 1 of 4

Desc ROUND TO THOUSANDS PLACE

Go To

1.	The item calls for an understanding of arithmetic operations -- including estimation -- on whole numbers, decimals, fractions or percents.	?
2.	The item calls for a thorough understanding of basic-level arithmetic operations -- an understanding sufficient for problem solving in practical situations.	?
3.	The item calls for understanding the connections among any of the following: fractions, percents, decimals.	?
4.	The item calls for using fundamental algebraic concepts in problem solving.	?
5.	The item calls for understanding of the connection between algebra and functions.	?
6.	The item involves geometric concepts.	?

Continue



Block 3 Item 1 ID M049901 Item 1 of 205 Page 3 of 4

Desc ROUND TO THOUSANDS PLACE Go To

14. The item involves statistics or probability. ?

15. The item calls for conceptual understanding or procedural understanding. ?

16. The item calls for applying mathematical concepts and procedures to complex problems. ?

17. The item calls for reaching beyond the recognition, identification, and application of mathematical rules to generalize and synthesize concepts and principles. ?

18. The item calls for comparing and contrasting mathematical ideas. ?

19. The item calls for generating one's own examples. ?

Continue

Block 3 Item 1 ID M049901 Item 1 of 205 Page 4 of 4

Desc ROUND TO THOUSANDS PLACE Go To

20. The item calls for probing of examples and counter examples in order to shape generalizations from which the student can develop models. ?

21. The item calls for the using number sense to consider the reasonableness of an answer. ?

22. The item calls for a written response. ?

Continue

Appendix F
Script for Judges' Training

Introduction and Welcome:

Purpose:

You are one of a group of mathematics educators who are helping us this week with one of our studies of the validity of the National Achievement of Educational Progress (NAEP). This study is trying to determine what kinds of mathematical knowledge the NAEP math test items are trying to measure.

Their Role:

Your role today is to match approximately 200 mathematics test items at a particular grade level (Grade 4, 8 or 12) to a set of categories/descriptors that describe various kinds of mathematical knowledge and skills. Each of you will work at one grade level. You will be asked to make the judgments based on your professional opinion and to NOT consult with others working on the same task. There are no right or wrong judgments.

We need to do several things before you start to work on your own:
1) Sign a nondisclosure form
2) Fill in a background questionnaire
3) Describe your task in detail
4) Demonstrate how you will use the binder of items and computer to make and record your judgments about each item.

Nondisclosure Forms:

Since you will be seeing the items used in the 1992 National Assessment and since some of these items will be used in future assessments, you will have to sign a "nondisclosure" form to agree not to communicate or use your knowledge of these items outside of the work you do here today. Please read the guidelines and, if you agree, sign the form.

A few minutes to read Guidelines and sign form

Any questions?

Background Information Questionnaire:

Before we describe your task in more detail, we want to obtain some background information about your mathematics and teaching experience. This information will be kept strictly confidential. We will use it to compile a

general description of the group of mathematics educators who acted as judges in the project. Please take a few minutes now to complete the questionnaire.

A few minutes to complete background questionnaire

The Materials:

Folders containing test items:
We have given each of you a binder (show a binder) that contains all of the items you will judge. Your name and the grade level of the items you will be judging are on the cover of the binder. You had to "sign in" to receive the binder; before you leave today you will have to "sign out" and return the binder to one of us. You cannot take the binder away from the Center. Your job is to decide which of a number of descriptions of mathematics content and skills characterize each of these items.
Each binder of items is divided into 14 "blocks" (show). At each grade level, three of you have the blocks running from Block 3 to 16, and three of you from Block 16 back to 3. So we have six teachers working on each grade level. Regardless of the order of blocks in your binder, you will work through the items, page by page, in the order that they occur in your binder. The pages in your binder are numbered from 1 to whatever. You will be able to tag and go back to items, but make sure that you do all items.

The "Descriptors"/Classification Form:
You have a paper copy of the full set of categories or "descriptors" that will be used to classify items for each grade level (there is a different set of categories for each grade level). We have computerized these forms and you will enter your judgments about items directly into one of the computers that you see set up here. (SHOW SAMPLE COMPUTER SCREEN ON PROJECTOR) As you can see for any one grade level, it takes a number of computer screens to display all of the descriptors to which an item might be matched. The number of "pages" on the top right-hand corner of the screen indicates which of the total set of descriptor screens you are on.

Take a few minutes to read the entire set of descriptors FOR THE GRADE LEVEL AT WHICH YOU WILL BE WORKING.

IMPORTANT:
You will notice that there is a degree of overlap among some of the descriptors, but we want you to check any descriptor that, in **your opinion**, describes a particular item. In some cases you may check a large number of the descriptors; in the case of other items, you may check only a few descriptors. You will also notice that some of the descriptors are ambiguous and their meaning is not defined. The descriptors come from actual text that we had to stick to verbatim; we have reordered the descriptors to ease the judgment task where possible. We want each of you to interpret the categories based on YOUR experience/impression of the mathematics curriculum and students at

the grade level for which the test was designed. You will be able to indicate when you are uncertain about any judgment that you make.

Let's do an example for each grade level.

Refer to their printed "logging on" instructions and demonstrate for them how to get into a 4th grade "stack"

Show a 4th grade item on the flip chart. Ask them to look at their printed version of the computer screens and check with a pencil the descriptors that apply.

Refer to their printed instructions for selecting descriptors on the computer. Demonstrate how one would use the mouse to select particular descriptors that this item fits
Also demonstrate how to indicate if you are unsure.
Demonstrate what happens if you do and don't select a descriptor which has subcategories. Mention that they can keep the paper copy of all the descriptors for their grade level to remind them of all of the possible options.

Then demonstrate the more general features of the computer screen:
Refer to their printed set of instructions for using the general features of the computer screen:

Block Number
Item number (verification required first time you come to it)
Description (of item)
GO TO (to go back and forth between item numbers)
Arrows (to go back and forth between the screens that contain the descriptors)
Continue (to move on to the next screen/page of descriptors)

Then do an 8th grade example (test item on flip chart, demonstrate again how to "log on" and get into a stack), and a 12th grade example.

Be clear that the decisions about what descriptors any item fits will vary from judge to judge. We are relying on their individual professional judgments. They should not discuss any item with other judges.

Any questions?

We will now get each of you started at the computer that has your name on it. We suggest that you work through 2 or 3 blocks of items at a time and then take a short break (refreshments will be available all day). Just mark the item/block where you finished and leave the computer screen as it is. Lunch will be available at noon so you can take a break at that time to eat and resume as soon as you can.

We are giving each of you a notepad and a pad of post-its. If you are having trouble with an item, you could mark it with a post-it (also, feel free to write on the items) and come back to it later (using the GoTo feature on the screen).

At the end of the day (or whenever you have completed the task for all items in your binder), you should leave the computer ON at the screen where you finished, return your binder to the person monitoring, and sign out. At that point we will ask you to give us written comments about any aspect of the activity that struck you as difficult, strange or unclear. In addition, we will ask to interview some of you about your experience with this activity.

Transition...

Now, put away any paper copy of the categories that do not apply to the grade level you will work on.

We will have someone monitoring you all day. If you have any questions, tell that person and he/she will come and get one of us (Brenda, Leigh, John) if necessary.

John, Leigh and I will get individuals started and monitor them as they do the first few items.

Set of materials to give each judge

1) Binder with judge's name and grade level on it
2) Paper copy of expanded descriptor screens for ALL three grade levels
3) Pad of paper, pen and pencil
4) Pad of post-its
5) Non-disclosure form and guidelines
6) Background questionnaire
7) Instructions for logging on, making selections, general features of the computer screen

INSTRUCTIONS FOR "LOGGING ON" TO YOUR STACK OF FORMS

Turn the machine on and wait until your name appears.
Use the mouse to move the cursor (hand or arrow) to the icon (stack) that has your name on it. Then double click the mouse button.

Wait a few seconds and the first screen of descriptors for the first item will appear.

INSTRUCTIONS FOR SELECTING A DESCRIPTOR

Use the mouse to move the cursor (hand or arrow) onto the number of the descriptor you want to select. Click the mouse button once.
The box around the number you selected will turn black.

If you want to undo a selection, just select it again and the box will return to its original "clear" state.

If you are unsure of any selection (or any descriptor/category that you have not selected), move the cursor to the box with the question mark in it beside the descriptor and click once to make that box turn black.

Once you have finished with one screen of descriptors, move the floating arrow to the "Continue" box in the lower right-hand corner and click the mouse button once. This will move you to the next "page"/screen of categories.

INSTRUCTIONS FOR GENERAL FEATURES ON COMPUTER SCREEN

The Block and Item number refer to the item that you are currently working on. As you move on to the next item, the item number changes and you will be asked to verify that this is the item you are considering in your binder.

You can use the arrow to the left of the "Go To" box to go back to an item you have already judged and change some selections for that item. When you click on the left hand GoTo arrow, a screen will pop up asking you to enter the number of the block you want to go back to. Once you type in a block number another screen will pop up asking you which item in that block you want to go to. When you have made your revision, you can then use the arrow to the right of the "Go To" box to go forward to the item you were on before you went back.

Once you have been through the pages of descriptors for any one item (2 pages of descriptors at 4th grade, 4 pages at 8th and 12th grade), you can click on the arrows in the upper right-hand corner to move back and forth between the pages for that item, in order to undo or add a selection.

Appendix G
Report of Judges' Oral and Written Comments*

Grade 4 Form

General Remarks

Five reviewers of the Grade 4 form made written comments. Only one reviewer was interviewed.

Two reviewers noted that they generally enjoyed the review process. One wrote "(o)verall it was fun," and the other expressed gratitude "for all the good food and an interesting day." Another reviewer wrote that "(i)t was interesting and fun for awhile (but) eventually became tedious and tiresome."

The two reviewers who said they enjoyed the process also noted both positive and negative reactions to the HyperCard program. One said simply that "(t)he stack worked well -- (but) I would have designed the card a little differently." The other said that "the powerbook was more interesting than just paper and pencil." In the interview she commented that this was the first time she had used a mouse. On the negative side she felt that the question mark was "too far from other responses" and suggested that "all responses would be easier to use to the right of the descriptions." She also noted that "(a)bout 1% of the time (17/176) when I hit the continue button after #4 it went to the next item rather than p. 2."

Several reviewers expressed doubts about the quality of the test items. One reviewer was particularly concerned about the ability of the items to accurately measure the math skills of LEP students. She said that she would like to see a style of testing closer to that of the new CAP where students would show more of their work and examples on the test and that these would be part of assessment. Three other reviewers commented on the lack of complexity and/or problem-solving in the test items. As one put it "the fourth grade test was fairly basic -- in fact, alarmingly simple with little opportunity to think or problem solve. I hope this doesn't represent the national state of 4th grade math."

The reviewer who was interviewed said that she tried to use multiple points of view, both her own and that of students when classifying the problems.

Comments on descriptors

Two reviewers wrote comments about the descriptors. These comments are marked by brackets ( ) below. Unbracketed comments are from the one reviewer who was interviewed.

---

* This report was prepared by Regie Stites who also conducted the interviews.

Descriptor 1
She tended to use the question mark with 1a and 1b because she saw a link between them as different phases in a unified process

Descriptor 2
She wondered what the word "some" meant in descriptor 2a.

Descriptors 3 and 4
She had trouble seeing the basic difference between 3 and 4 and felt that they were overlapping.

Descriptor 5
I asked about her understanding of the term "complex" and she said that she understood it to refer to multi-step problems.
She also was uncertain about the meaning of "non-routine" and commented that she considered graphing to be a "non-routine" activity because students rarely did graphs.

Descriptor 6
She felt that this descriptor was difficult to apply because of the term "real world.
She commented that "just about everything" could be seen as a real world problem.
She also noted that she had trouble making the distinctions called for in the subcategories.
(Two reviewers wrote comments on descriptors 4-6. One noted that "4 lacks reference to real world, but 6c repeats 5 without 'integration," and how to include a complex routine problem." He went on to note that "simple translation interpretation comprehension/application is not problem solving." A second reviewer wrote that for some problems "(i)t was not clear what the goal of the activity was as many items were too close were too close in content or intent while others like 'solving a simple real-world problem vs. solving a [routine] real-world problem' drove me wild -- purely a semantics problem. What is the difference? What is real-world?")

Descriptor 7
She used this descriptor a lot because she felt that it was generally applicable and used it even more at the end of the day.

Descriptor 8
She was bothered by the use of "how" vs "why" in 8 b and c respectively. She did not see a difference in applying these and therefore she tended to use only 8b.
She also felt that 8c and d were repetitions of 8a and b and was bothered by the use of "clear and concise" in 8d.
(Another reviewer wrote "(n)one of the items seemed complex and I would have liked more choices under 8 written response. All items seemed to require some basic problem solving and math understanding."

Suggestions for additions

Two reviewers made suggestions for additional descriptors in their written comments. One noted that "given these particular descriptors" it was "very difficult ... to talk about measurement or reading/rendering graphic representations." Similarly, a second reviewer wrote that "there were no responses to deal (adequately) with measurement or identifying patterns," and went on to note that she felt she was dealing with problems that contained such features by "squeezing them in where they didn't quite fit."
The first of these reviewers also noted a "question (as to) whether indeed there is any descriptor there for spatial reasoning."

Comments on the NAEP categories

None of the written comments referred directly to the NAEP classification scheme, but several reviewers made comments which could be interpreted as generally critical of the language used in the scheme. One reviewer wrote that she "found the ambiguity of the terms very bothersome at first." She went on to say that by "going over possible meanings several dozen times, (she) decided what they meant to (her), assigned definitions and proceeded on that basis." Another reviewer noted that the lack of "problem solving hierarchy ... in the descriptors was a problem." A third reviewer noted that "(a)s (she) went along, the stipulations to a problem seemed more to make sense. At times it was confusing to pick one item over another. At other times they seemed to overlap."

Grade 8 Form

General Remarks

Six reviewers of the Grade form wrote comments on the process. In addition I interviewed three of these reviewers.

All three interviewed reviewers were generally happy with the review process. One was especially pleased with the HyperCard format, but suggested that the "continue" button might be changed to a return function in order to reduce movement of the mouse.

Written comments on the overall process were also generally favorable. One reviewer wrote "(t)he staff and their treatment of us was wonderful."

When asked during the interview, two reviewers said that they tried to analyze the problems from the students' viewpoint. The third said that she used her own point of view. In her written comments she noted that one "... judgment call I kept running into was whether to look at the written answer from a student's point of view [what was specifically asked for] or an evaluator's point of view [what makes a good answer]"

One reviewer who was not interviewed made written comments that seem to indicate that she was analyzing the problems from a student perspective. For example, she wrote in reference to descriptor 15 the she felt that "to solve many of the problems the student would need to have a conceptual understanding."

One of the interviewed reviewer noted in her written comments that in "classifying some items, a lot would depend on the type of test taker a person is -- many people estimate all answers rather than working them out or generate when they aren't required."

All three interviewed reviewers said that they made frequent use of the question mark. Two said that they used the question mark to indicate their lack of clarity as to the meaning of the descriptor and to its fit to the problem. The other said that he used the question mark when he felt that the descriptor applied to an "implicit" rather than an "explicit" feature of the problem.

Comments on descriptors

Two reviewers who were not interviewed wrote general comments on the descriptors. Both noted the ambiguity of the descriptors. One wrote "(m)any of the descriptors were very ambiguous and subjective -- 'unique,' 'complex,' 'fundamental,' 'mathematical ideas' are open to interpretation." The other wrote "I tried to be consistent with the descriptors as they applied to different questions but I'm afraid I did not always accomplish (the) task [especially with

the earlier sections]. I felt that some descriptors were intentionally ambiguous."

The first of these reviewers also wrote "I found none/few of these items met some of the criteria: #11, 16, 17, 18." Another reviewer also wrote that descriptors 11 and 17 "seemed to apply to very few of the problems."

The following comments are based on the interviews except where noted otherwise.

Descriptor 1
One reviewer used this descriptor frequently because he felt that it applied to most of the problems.

Descriptors 1 and 2
One reviewer had difficulty distinguishing between descriptors 1 and 2, especially in the early stages of the review. She eventually settled on the idea of estimation as the discriminating factor, but noted that there was no option for basic understanding without estimation

Descriptor 2
One reviewer felt that this descriptor "was critical."

Descriptor 3
One reviewer commented that this descriptor represented an important skill that was not often reflected in the test items.
Another reviewer felt that this descriptor was only partially applicable in some cases.
(One reviewer who was not interviewed wrote that "I found item #3 was very confusing -- I was unsure what relationships they were referring to." Another wrote "(o)n descriptor 3, if any two items were related, I indicated that descriptor applied."

Descriptor 4
One reviewer mentioned that he did not see a lot of items which called for the application of this descriptor. He did mark it in cases where use of algebraic concepts were not necessary but could have been used.

Descriptor 5
One reviewer felt that this descriptor was difficult to interpret because the connection between algebra and functions was not clear. As a result the reviewer did not use this descriptor often.

Descriptor 6
One reviewer noted that he used this descriptor a lot, but he also said that he was surprised that the geometry in the items was all formal and would have liked to see more informal geometry.

Descriptor 7
All three reviewers commented on this item. One felt that meaning of "spatial relationships" was unclear and used his own qualification to make sense of it. Both other reviewers said that they often used this descriptor. One said that she used it when reading a graph was required and the other said that he used it for problems that involved visualization.

Descriptor 8
One reviewer commented that he seldom used this descriptor.

Descriptor 9
One reviewer commented that this descriptor seemed to apply to almost every problem, but that she tried to limit its use to more abstract problems which required stopping and thinking about how to solve them.
Another reviewer commented that this descriptor did not seem to apply to many problems because there were not a lot of different strategies that could be used to solve the test items.

Descriptor 10
One reviewer commented that he did not often use this descriptor because he did not see problems that called for the use of calculators.
Another reviewer had wondered whether a ruler could be considered a "technological tool" and had decided that it could be.

Descriptor 12
One reviewer commented that he did not use this descriptor (often).

Descriptor 13
One reviewer commented that there were some "working with data" items that did not involve the skills described in either 13a or b.

Descriptor 14
One reviewer felt that the subcategories under this descriptor were too limited and would have liked to add geometry here.

Descriptor 15
All three reviewers commented on this descriptor. All felt that it was very generally applicable. Two reviewers noted that the reference to "conceptual or procedural understanding" made the descriptor nearly universally applicable.
(One of the reviewers who was not interviewed noted that descriptor 15 was "used on most of the problems" because "to solve many of the problems the student would need to have a conceptual understanding.")

Descriptors 15 and 16
One reviewer felt that these two were hard to differentiate.

Descriptor 16
One reviewer said that she did not use this descriptor very often because of the term "complex," which she interpreted as eliminating problems which only required one or two steps.
I asked another reviewer about the use of the term "complex" in this descriptor and he agreed that it was somewhat difficult to interpret though he did not seem to have had problems with this in the rating.

Descriptors 16, 17, and 18
One reviewer did not use these three descriptors often because he felt that they applied to few problems.

Descriptor 17
One reviewer said that they did not use this descriptor often because there were very few complex problems

Descriptors 18 and 19
One reviewer said that these two descriptors seemed to be the same.

Descriptor 19
One reviewer commented that "kids don't do much of this."
Another reviewer used the ? with this descriptor because she saw the possibility of different strategies to solving some items.

Descriptor 20
One reviewer felt that there were few situations in the problems that could be used to develop models.
Another reviewer felt this descriptor was related to higher level problems (beyond the 8th grade level).

Descriptor 21
All three reviewers commented on this descriptor. One pointed out a typo (the addition of "the"). One said that he used a lot and another said that he seldom did. The third explained that its application may depend on the kind of test taker you are, therefore she only chose it when the problem really required it.

Descriptor 22
All three reviewers commented on this item. One said that he used this descriptor and its subcategories, especially b-d.
Another said that 22a and e were both "for higher level stuff."
Another wondered whether answering by filling in a circle would constitute an "example."
(In her written comments the last reviewer noted "(t)he problems that required written answers were also hard to classify -- there was no classification for it being just a written answer -- were they counted as examples?")

Suggestions for additions

There were essentially six suggestions for additional descriptors. One reviewer would have liked to have seen additional descriptors for statistics and probability and more geometric descriptors. Another felt that a descriptor for simple "measuring;" one for "graphing;" one for "number sense in general" that would not be as specific as 21; and one for "patterns" for understanding of a repeated pattern should be added.

Comments on NAEP categories

One reviewer said that her initial response to the NAEP classification schema was that it seemed that only difference between the basic and proficient levels was that in the latter you needed to write out (show work on) the answers. She also wondered about problems that may have two different correct answers and how this would be evaluated.

Grade 12 Form

General Remarks

Five of the Grade 12 form reviewers wrote comments on the process. I interviewed four of these reviewers.

With regard to the review process, the reviewers had generally favorable reactions and noted that the computer made the task much easier and that the other materials were also okay. Some felt that working over two days was better than doing all the work in a single eight hour period, although one reviewer did say that work at a stretch helped her to focus.

In discussing the manner in which they approached the task one reviewer said that in the beginning it was slow because he would have to think for a long time to get the ideas about the descriptors clear. Another said that he worked through a process of eliminating descriptors that clearly did not apply and then considered which of the remaining might. He noted that he would like to have worked out a solution to each problem before deciding where to place it but that this was not possible within the time given.

Three of the four reviewers said that they tried to see multiple possible solutions to problems and used their own as well as students' point of view. Two reviewer used the question mark to indicate the possibility of multiple solutions. One reviewer said that he used his own judgment in rating items and was not thinking of how students would respond.

In the written comments one reviewer noted that "(s)ome of the problems could be solved by a variety of methods and the process was not immediately clear in some of them. The ability of students to recognize concepts and work with them depends on their depth of understanding the material."

One reviewer said that working alone was a problem mostly because he would have liked to have had someone to discuss his interpretations of the meaning of the descriptors with and because he needed to do a lot of thinking about some of the descriptors.

One reviewer suggested that more information be provided about the descriptors before the task and described how he came to change his understanding and application of the descriptors as he did the items He also suggested that more stress needs to given to idea that raters can decide for themselves what the descriptors mean.

Three reviewers wrote comments on the difficulty of classifying the problems using the descriptors. One wrote that "(a)t times it was difficult to 'fit' a problem into the language of the descriptors. In some cases, there was some overlap between descriptors. In other cases, there weren't any descriptors available to accurately describe a problem." Another noted that "(a)s is

probably the norm in most cases of categorizing, I found either that the categories were sometimes ambiguous, not enough differentiation between two categories, or the category I thought appropriate was nowhere to be found. After making my own decisions on how to interpret each one, it went much faster."

One reviewer suggested that "(p)erhaps like problems should be grouped together -- making comparing and contrasting them more accurate."

Comments on descriptors

The following comments are based on interviews with four reviewers. None of the reviewer made written comments about the descriptors other than those already noted above.

Descriptor 1
Two reviewers commented on descriptor 1 subcategories. One said that she used the ? to indicate her lack of precise understanding of the meaning of the terms "justifying" (in 1e) and the distinction between "using" and "understanding" (in 1a and 1c -- also in 2a and 2b). She said that she applied 1e rarely and then only to problems that called for "showing something." She also noted that in cases where the problem called for actually measuring something she had to choose between 1a and 11 and used ? to indicate the difficulty in making this choice.
Another reviewer found 1f to be problematic for him because he did not know whether it implied that a pattern was already provided or if the student needed to generate a pattern.

Descriptor 2
One reviewer commented that 2c was made difficult by the addition of the term "polynomials." She felt that this was too specific and therefore generally ignored this part of the descriptor when applying it. She did this because she saw no other place to put algebraic operations. She also had trouble differentiating between 2a and 2b (see above).

Descriptor 3
One reviewer said that he did not often use this descriptor because he did not see many problems to place there.
Another reviewer did not like the term "concept" (in the subcategories) and would have preferred the term "properties."

Descriptor 4
One reviewer said that he liked this descriptor and felt that it was clear.

Descriptor 5
One reviewer said that he did not use this descriptor very much because he found the terms "continuous" and "discrete" difficult to interpret.

Descriptor 6
Two reviewers commented on this descriptor. Both said that they interpreted it to apply very broadly. One said that he used it for all items that he had difficulty classifying that were not real world problems. The other said that she used it often for any problems that called for interpreting graphs. She also said that her reading of "recognizing relationships" was very broad.

Descriptor 7
Two reviewers commented that this descriptor was difficult to interpret. One said that not many problems called for "models." The other found the difficulty in the reference to "counterexamples." He felt that this was something that only college level students should/could be expected to do.

Descriptor 8
One reviewer noted that she used this descriptor even if the problem was not clearly a real world application.

Descriptors 8, 9, and 10
Three reviewers noted problems with the terminology of these three descriptors. Two were troubled by the reference to "real world problems." One said that it was especially hard to differentiate real world from non real world problems when one had to also consider the other criteria in the descriptors. He said that if the "real world" qualification had been left off he would have had less trouble applying these descriptors (8 & 9).
Another reviewer said that she was not certain about what made a problem a "real world" problem. She wondered aloud whether "computation" by itself was a real world problem.
Two reviewers had trouble with the term "estimation" in these descriptors. One reviewer said that she found that she often vacillated in choosing 8 or 9, because the difference between "estimation to verify" and "estimation to determine" was a difficult one to make in "real world problems."
Another reviewer also had difficulty with the term "estimation" and therefore he did not use these descriptors often. This avoidance was also partly due to his indecision about what "real world" meant.

Descriptor 10
One reviewer noted that she saw few problems that called for "judging or defending the reasonableness of answers" and therefore did not use this descriptor often.

Descriptors 10 and 11
One reviewer said that he liked these two descriptors .

Descriptor 11
All four reviewers said that they liked and/or often used this descriptor. All interpreted the descriptor in a very general way and therefore used it often.
Two reviewers noted that almost all problems would involve some sort of "procedural" or "conceptual knowledge" and therefore they frequently used this descriptor.

One reviewer said that any problem that involved addition or subtraction would go here because to do these operations you need to understand them.
Another said that she felt that some questions involved number theory and since she did not see a specific descriptor for this she interpreted it as "conceptual knowledge."
Another felt that some pattern recognition in the problems did not fit in the geometric or other general areas and therefore she would put it under the general descriptor (11) and use it as a "how to" descriptor.
One reviewer did not feel comfortable with the language of "procedure" in this descriptor. He felt that just about every item involved some level of procedural knowledge and felt therefore that this was too general a descriptor. He also had difficulty interpreting the meaning of "procedural knowledge or conceptual knowledge" and wondered whether rounding of fractions could be considered an example of this. This is a case of something that he would have liked to discuss with others or to have had more training on. He felt that he was more discriminating in using this descriptor at the end and suggested that some words be added to specify level of procedural knowledge. He also suggested that raters be asked to read descriptors 11-13 and to think about them before they begin rating so that they have a clearer idea of how to apply them

Descriptor 12
Two reviewers commented on the use of the term "complex" in this descriptor. One said that he was not sure what was meant by a "complex" problem since from his point of view a problem may be simple but still be complex from the students' viewpoint.
The other said that he only used this descriptor when a problem required a number of smaller steps to be taken to find a solution and that the descriptor was used more at the end of the day than at the beginning. He thought this was because his understanding of the descriptors had become clearer by that point

Descriptors 12 and 13
Two reviewers commented on the term "integrating/integration" in descriptors 12 and 13. One said that he did not often use these descriptors because he was not clear about the meaning of "integrating."
The other reviewer felt that 12 was easy to apply but had difficulty with 13.

Descriptor 13
Two reviewers had difficulty with the phrase "synthesis of ideas" in 13. He said that the few time he applied 13 he marked it with the ? because he was unsure of the meaning of "synthesis."
Another reviewer said that he assumed that "synthesis of ideas" referred to problems "that would draw from different areas" or that would entail "contrasts/combinations of ideas."

Descriptor 14
One reviewer said that he did use this descriptor often but not the sub-categories because the "and" in the subcategories should have been "and/or."

Suggestions for additions

Two reviewers suggested that a descriptor be added for low level computation. One described the needed descriptor as applying to problems that called for computation with "no reasoning behind it." The other said that he would have liked to have chosen "none of the above" at times and was missing a lower level of computation that could be worded as "straight computation," or "common sense," or "straight procedure."

One reviewer said that she would have liked to see a descriptors for "recognizing patterns," for general problem solving (geometric, algebraic, etc.), and for measurement skills.

Comments on NAEP categories

Three reviewers offered general reactions to the NAEP classifications. One-said that it would have helped if he had had prior knowledge of where descriptors fell in the basic to advanced continuum.

Another saw the transfer of research to the classroom as a general problem and talked about the "new math" and its failure to gain acceptance by teachers. His general opinion was that teachers need "retraining" in order to be able to make use of the latest research and he was skeptical about the NAEP rating scheme for this reason.

One said that she was not happy with the use of "reasoning skills" as a descriptor for the basic level since she felt that most students are not capable of this.

Appendix H
Number of Descriptors From Each Level Mapped to Each Test Item

GRADE 4 ITEMS ORDERED BY TOTAL NUMBER OF DESCRIPTORS MAPPED

			Number of Descriptors Consistently Mapped			
NAEPID	BLOCK	ITEM	TOTAL	BASIC	PROFICIENT	ADVANCED
M061905	10	5	6.00	.00	4.00	2.00
M049001	15	10	6.00	.00	4.00	2.00
M022301	5	5	5.00	.00	5.00	.00
M023401	5	17	5.00	1.00	4.00	.00
M045401	7	10	5.00	.00	4.00	1.00
M061906	10	6	5.00	.00	4.00	1.00
M044401	14	10	5.00	.00	4.00	1.00
M032901	16	10	5.00	1.00	4.00	.00
M039201	3	3	4.00	1.00	3.00	.00
M039801	3	9	4.00	.00	4.00	.00
M039901	3	10	4.00	.00	4.00	.00
M040001	3	11	4.00	.00	4.00	.00
M017801	4	5	4.00	.00	4.00	.00
M017901	4	6	4.00	.00	4.00	.00
M018601	4	13	4.00	.00	4.00	.00
M021901	5	1	4.00	.00	4.00	.00
M022001	5	2	4.00	1.00	3.00	.00
M023201	5	15	4.00	.00	4.00	.00
M045001	7	6	4.00	.00	4.00	.00
M045101	7	7	4.00	.00	4.00	.00
M045201	7	8	4.00	1.00	3.00	.00
M010531	8	6	4.00	.00	4.00	.00
M040701	9	5	4.00	.00	4.00	.00
M041201	9	10	4.00	.00	3.00	1.00
M061901	10	1	4.00	.00	2.00	2.00
M043301	13	8	4.00	.00	4.00	.00
M043501	13	12	4.00	.00	3.00	1.00
M044101	14	6	4.00	.00	4.00	.00
M048601	15	6	4.00	.00	4.00	.00
M032201	16	3	4.00	1.00	3.00	.00
M032401	16	5	4.00	2.00	2.00	.00
M031401	16	14	4.00	1.00	3.00	.00
M031801	16	19	4.00	.00	4.00	.00
M039401	3	5	3.00	.00	2.00	1.00
M040201	3	13	3.00	.00	3.00	.00
M018501	4	12	3.00	.00	3.00	.00
M022701	5	9	3.00	.00	3.00	.00
M022901	5	12	3.00	1.00	2.00	.00
M023001	5	13	3.00	.00	3.00	.00
M023301	5	16	3.00	.00	3.00	.00
M020701	6	11	3.00	.00	2.00	1.00
M045301	7	9	3.00	.00	2.00	1.00
M010331	8	4	3.00	.00	3.00	.00
M010431	8	5	3.00	.00	3.00	.00
M040601	9	4	3.00	.00	3.00	.00

M046301	11	4	3.00	.00	3.00	.00
M046801	11	9	3.00	.00	3.00	.00
M047101	11	12	3.00	.00	3.00	.00
M047201	11	13	3.00	.00	3.00	.00
M041501	12	3	3.00	.00	3.00	.00
M041901	12	7	3.00	.00	3.00	.00
M042001	12	8	3.00	.00	3.00	.00
M042002	12	9	3.00	.00	3.00	.00
M042003	12	10	3.00	.00	3.00	.00
M042501	12	12	3.00	.00	3.00	.00
M043701	14	2	3.00	.00	3.00	.00
M044201	14	7	3.00	1.00	2.00	.00
M044301	14	9	3.00	.00	3.00	.00
M048701	15	7	3.00	.00	3.00	.00
M031501	16	16	3.00	1.00	2.00	.00
M031901	16	20	3.00	.00	3.00	.00
M039601	3	7	2.00	.00	2.00	.00
M039701	3	8	2.00	.00	2.00	.00
M017401	4	1	2.00	1.00	1.00	.00
M018101	4	8	2.00	.00	2.00	.00
M018201	4	9	2.00	1.00	1.00	.00
M018301	4	10	2.00	.00	2.00	.00
M018401	4	11	2.00	.00	2.00	.00
M022601	5	8	2.00	.00	2.00	.00
M022802	5	11	2.00	.00	2.00	.00
M019701	6	1	2.00	1.00	1.00	.00
M020201	6	6	2.00	.00	1.00	1.00
M020401	6	8	2.00	.00	2.00	.00
M020501	6	9	2.00	.00	2.00	.00
M010731	8	8	2.00	1.00	1.00	.00
M010831	8	9	2.00	1.00	1.00	.00
M010931	8	10	2.00	.00	2.00	.00
N240031	8	14	2.00	.00	2.00	.00
M040501	9	3	2.00	.00	2.00	.00
M040901	9	7	2.00	.00	2.00	.00
M041001	9	8	2.00	.00	2.00	.00
M041101	9	9	2.00	.00	2.00	.00
M046601	11	7	2.00	.00	2.00	.00
M046901	11	10	2.00	.00	2.00	.00
M047301	11	14	2.00	.00	2.00	.00
M041701	12	5	2.00	.00	2.00	.00
M042401	12	11	2.00	.00	2.00	.00
M042801	13	3	2.00	1.00	1.00	.00
M042901	13	4	2.00	1.00	1.00	.00
M043801	14	3	2.00	.00	2.00	.00
M048201	15	2	2.00	.00	2.00	.00
M048901	15	9	2.00	.00	2.00	.00
M032301	16	4	2.00	1.00	1.00	.00
M032701	16	8	2.00	1.00	1.00	.00
M032801	16	9	2.00	.00	2.00	.00
M031201	16	12	2.00	1.00	1.00	.00
M031301	16	13	2.00	1.00	1.00	.00
M031402	16	15	2.00	.00	2.00	.00
M031701	16	18	2.00	.00	2.00	.00

M039101	3	2	1.00	.00	1.00	.00
M039501	3	6	1.00	.00	1.00	.00
M040101	3	12	1.00	.00	1.00	.00
M017501	4	2	1.00	.00	1.00	.00
M017601	4	3	1.00	.00	1.00	.00
M018001	4	7	1.00	.00	1.00	.00
M018701	4	14	1.00	.00	1.00	.00
M022101	5	3	1.00	.00	1.00	.00
M022201	5	4	1.00	.00	1.00	.00
M022401	5	6	1.00	.00	1.00	.00
M022501	5	7	1.00	.00	1.00	.00
M022801	5	10	1.00	.00	1.00	.00
M023101	5	14	1.00	.00	1.00	.00
M020101	6	5	1.00	.00	1.00	.00
M020301	6	7	1.00	.00	1.00	.00
N277903	6	10	1.00	.00	1.00	.00
M044501	7	1	1.00	.00	1.00	.00
M044601	7	2	1.00	.00	1.00	.00
M044701	7	3	1.00	.00	1.00	.00
M044901	7	5	1.00	1.00	.00	.00
N214331	8	1	1.00	.00	1.00	.00
M010131	8	2	1.00	.00	1.00	.00
N202831	8	12	1.00	.00	1.00	.00
M011231	8	15	1.00	.00	1.00	.00
M040301	9	1	1.00	.00	1.00	.00
M040401	9	2	1.00	.00	1.00	.00
M040402	9	2	1.00	.00	1.00	.00
M040403	9	2	1.00	.00	1.00	.00
M040801	9	6	1.00	.00	1.00	.00
M046001	11	1	1.00	.00	1.00	.00
M046701	11	8	1.00	.00	1.00	.00
M047001	11	11	1.00	.00	1.00	.00
M041301	12	1	1.00	.00	1.00	.00
M041601	12	4	1.00	.00	1.00	.00
M042601	13	1	1.00	.00	1.00	.00
M042701	13	2	1.00	.00	1.00	.00
M043201	13	7	1.00	.00	.00	1.00
M043401	13	9	1.00	.00	1.00	.00
M043402	13	10	1.00	.00	1.00	.00
M043403	13	11	1.00	.00	1.00	.00
M043601	14	1	1.00	.00	1.00	.00
M043901	14	4	1.00	.00	1.00	.00
M044202	14	8	1.00	.00	1.00	.00
M048101	15	1	1.00	.00	1.00	.00
M048501	15	5	1.00	.00	1.00	.00
M032001	16	1	1.00	1.00	.00	.00
M032101	16	2	1.00	1.00	.00	.00
M032501	16	6	1.00	1.00	.00	.00
M032601	16	7	1.00	1.00	.00	.00
M031101	16	11	1.00	.00	1.00	.00
M031601	16	17	1.00	1.00	.00	.00
M039001	3	1	.00	.00	.00	.00
M039301	3	4	.00	.00	.00	.00
M017701	4	4	.00	.00	.00	.00

M019801	6	2	.00	.00	.00	.00
M019901	6	3	.00	.00	.00	.00
M020001	6	4	.00	.00	.00	.00
M044801	7	4	.00	.00	.00	.00
M010231	8	3	.00	.00	.00	.00
M010631	8	7	.00	.00	.00	.00
N250231	8	11	.00	.00	.00	.00
M011131	8	13	.00	.00	.00	.00
M061902	10	2	.00	.00	.00	.00
M061903	10	3	.00	.00	.00	.00
M061904	10	4	.00	.00	.00	.00
M046101	11	2	.00	.00	.00	.00
M046201	11	3	.00	.00	.00	.00
M046401	11	5	.00	.00	.00	.00
M046501	11	6	.00	.00	.00	.00
M047401	11	15	.00	.00	.00	.00
M047501	11	16	.00	.00	.00	.00
M041401	12	2	.00	.00	.00	.00
M041801	12	6	.00	.00	.00	.00
M043001	13	5	.00	.00	.00	.00
M043101	13	6	.00	.00	.00	.00
M044001	14	5	.00	.00	.00	.00
M048301	15	3	.00	.00	.00	.00
M048401	15	4	.00	.00	.00	.00
M048801	15	8	.00	.00	.00	.00

GRADE 8 ITEMS ORDERED BY TOTAL NUMBER OF DESCRIPTORS MAPPED

			Number of Descriptors Consistently Mapped			
NAEPID	BLOCK	ITEM	TOTAL	BASIC	PROFICIENT	ADVANCED
M055101	14	5	11.00	5.00	4.00	2.00
M055501	14	9	10.00	3.00	6.00	1.00
M051101	3	13	9.00	2.00	6.00	1.00
M052201	13	11	9.00	3.00	6.00	.00
M053101	9	9	8.00	1.00	6.00	1.00
M061907	10	5	8.00	2.00	4.00	2.00
M061905	10	7	8.00	1.00	6.00	1.00
M054301	12	9	8.00	4.00	4.00	.00
M052001	13	9	8.00	4.00	2.00	2.00
M055401	14	8	8.00	4.00	3.00	1.00
M034001	16	21	8.00	4.00	3.00	1.00
M019201	4	19	7.00	4.00	3.00	.00
M019601	4	21	7.00	3.00	3.00	1.00
M012631	8	5	7.00	5.00	1.00	1.00
M053801	12	4	7.00	5.00	2.00	.00
M054001	12	6	7.00	4.00	3.00	.00
M051801	13	7	7.00	3.00	3.00	1.00
M054801	14	2	7.00	2.00	4.00	1.00
M055301	14	7	7.00	4.00	2.00	1.00
M032701	16	8	7.00	4.00	2.00	1.00
M022201	5	4	6.00	4.00	2.00	.00
M021001	6	12	6.00	3.00	3.00	.00

M013031	8	9	6.00	3.00	3.00	.00
M013231	8	11	6.00	3.00	3.00	.00
M055201	14	6	6.00	3.00	2.00	1.00
M049501	15	14	6.00	3.00	2.00	1.00
M034101	16	22	6.00	4.00	2.00	.00
M017801	4	5	5.00	2.00	2.00	1.00
M018301	4	10	5.00	2.00	2.00	1.00
M018601	4	13	5.00	2.00	2.00	1.00
M018801	4	15	5.00	3.00	2.00	.00
M021901	5	1	5.00	3.00	1.00	1.00
M022001	5	2	5.00	2.00	2.00	1.00
M022802	5	11	5.00	3.00	2.00	.00
M045901	7	13	5.00	2.00	3.00	.00
M012931	8	8	5.00	4.00	1.00	.00
M053001	9	8	5.00	2.00	3.00	.00
M061901	10	1	5.00	1.00	3.00	1.00
M061908	10	6	5.00	1.00	3.00	1.00
M046701	11	8	5.00	2.00	2.00	1.00
M047901	11	18	5.00	2.00	3.00	.00
M048001	11	19	5.00	2.00	2.00	1.00
M054101	12	7	5.00	3.00	2.00	.00
M054901	14	3	5.00	3.00	2.00	.00
M049301	15	12	5.00	3.00	1.00	1.00
M032301	16	4	5.00	3.00	1.00	1.00
M033001	16	11	5.00	4.00	1.00	.00
M050301	3	5	4.00	1.00	2.00	1.00
M017701	4	4	4.00	2.00	2.00	.00
M018001	4	7	4.00	2.00	2.00	.00
M018101	4	8	4.00	.00	3.00	1.00
M018501	4	12	4.00	2.00	1.00	1.00
M019001	4	17	4.00	1.00	2.00	1.00
M022301	5	5	4.00	2.00	1.00	1.00
M022501	5	7	4.00	2.00	2.00	.00
M023101	5	14	4.00	3.00	1.00	.00
M023201	5	15	4.00	2.00	2.00	.00
M023701	5	20	4.00	3.00	1.00	.00
M020901	6	11	4.00	2.00	2.00	.00
M021101	6	13	4.00	1.00	2.00	1.00
M021301	6	15	4.00	2.00	2.00	.00
M021302	6	16	4.00	2.00	2.00	.00
M044601	7	2	4.00	2.00	2.00	.00
M045301	7	9	4.00	2.00	2.00	.00
M012331	8	2	4.00	2.00	1.00	1.00
M012431	8	3	4.00	2.00	1.00	1.00
M011131	8	13	4.00	2.00	1.00	1.00
M013431	8	15	4.00	2.00	1.00	1.00
M052501	9	3	4.00	1.00	3.00	.00
M061903	10	2	4.00	2.00	2.00	.00
M061902	10	4	4.00	2.00	2.00	.00
M046601	11	7	4.00	1.00	3.00	.00
M053901	12	5	4.00	2.00	2.00	.00
M054201	12	8	4.00	2.00	1.00	1.00
M051601	13	5	4.00	2.00	1.00	1.00
M055001	14	4	4.00	2.00	1.00	1.00

M048201	15	2	4.00	2.00	2.00	.00
M048701	15	7	4.00	.00	3.00	1.00
M049401	15	13	4.00	3.00	1.00	.00
M049701	15	16	4.00	2.00	1.00	1.00
M049801	15	17	4.00	1.00	2.00	1.00
M032201	16	3	4.00	2.00	2.00	.00
M032601	16	7	4.00	1.00	2.00	1.00
M033401	16	15	4.00	3.00	1.00	.00
M033501	16	16	4.00	2.00	1.00	1.00
M033601	16	17	4.00	2.00	2.00	.00
M033801	16	19	4.00	2.00	1.00	1.00
M033901	16	20	4.00	2.00	2.00	.00
M050201	3	4	3.00	1.00	2.00	.00
M050202	3	4	3.00	1.00	2.00	.00
M050203	3	4	3.00	1.00	2.00	.00
M050204	3	4	3.00	1.00	2.00	.00
M050601	3	8	3.00	2.00	1.00	.00
M050701	3	9	3.00	2.00	1.00	.00
M050801	3	10	3.00	2.00	.00	1.00
M050901	3	11	3.00	2.00	1.00	.00
M051001	3	12	3.00	2.00	1.00	.00
M017901	4	6	3.00	.00	2.00	1.00
M018401	4	11	3.00	1.00	1.00	1.00
M018701	4	14	3.00	1.00	1.00	1.00
M019301	4	20	3.00	2.00	1.00	.00
M022101	5	3	3.00	2.00	1.00	.00
M022601	5	8	3.00	2.00	1.00	.00
M022701	5	9	3.00	2.00	1.00	.00
M022801	5	10	3.00	2.00	1.00	.00
M022901	5	12	3.00	1.00	1.00	1.00
M023001	5	13	3.00	1.00	1.00	1.00
M023301	5	16	3.00	1.00	2.00	.00
M023401	5	17	3.00	.00	2.00	1.00
M023501	5	18	3.00	1.00	2.00	.00
M023801	5	21	3.00	1.00	2.00	.00
M020201	6	6	3.00	1.00	2.00	.00
M020301	6	7	3.00	2.00	1.00	.00
M020801	6	10	3.00	2.00	1.00	.00
M044701	7	3	3.00	.00	3.00	.00
M045001	7	6	3.00	1.00	1.00	1.00
M045201	7	8	3.00	2.00	1.00	.00
M045601	7	10	3.00	2.00	1.00	.00
M012231	8	1	3.00	2.00	.00	1.00
M012531	8	4	3.00	1.00	1.00	1.00
M012731	8	6	3.00	2.00	1.00	.00
M013631	8	17	3.00	2.00	1.00	.00
M013731	8	18	3.00	2.00	1.00	.00
M052601	9	4	3.00	2.00	1.00	.00
M052701	9	5	3.00	1.00	2.00	.00
M052901	9	7	3.00	2.00	1.00	.00
M061904	10	3	3.00	1.00	2.00	.00
M046201	11	3	3.00	1.00	2.00	.00
M046301	11	4	3.00	1.00	1.00	1.00
M046401	11	5	3.00	1.00	1.00	1.00

M047201	11	13	3.00	1.00	2.00	.00
M047601	11	15	3.00	2.00	1.00	.00
M047701	11	16	3.00	2.00	1.00	.00
M051201	13	1	3.00	1.00	1.00	1.00
M051401	13	3	3.00	2.00	1.00	.00
M051901	13	8	3.00	2.00	1.00	.00
M052101	13	10	3.00	2.00	1.00	.00
M054701	14	1	3.00	2.00	1.00	.00
M049201	15	11	3.00	1.00	2.00	.00
M049601	15	15	3.00	2.00	.00	1.00
M032101	16	2	3.00	1.00	1.00	1.00
M032401	16	5	3.00	2.00	.00	1.00
M032501	16	6	3.00	2.00	.00	1.00
M032801	16	9	3.00	2.00	1.00	.00
M032901	16	10	3.00	2.00	1.00	.00
M033101	16	12	3.00	1.00	1.00	1.00
M033301	16	14	3.00	1.00	1.00	1.00
M033701	16	18	3.00	1.00	1.00	1.00
M049901	3	1	2.00	1.00	.00	1.00
M050001	3	2	2.00	1.00	1.00	.00
M050101	3	3	2.00	1.00	1.00	.00
M050501	3	7	2.00	1.00	1.00	.00
M017401	4	1	2.00	1.00	.00	1.00
M017501	4	2	2.00	.00	1.00	1.00
M018201	4	9	2.00	.00	1.00	1.00
M018901	4	16	2.00	1.00	1.00	.00
M019101	4	18	2.00	2.00	.00	.00
M022401	5	6	2.00	1.00	1.00	.00
M023601	5	19	2.00	2.00	.00	.00
M019901	6	3	2.00	1.00	1.00	.00
M020101	6	5	2.00	1.00	1.00	.00
M020401	6	8	2.00	2.00	.00	.00
M020501	6	9	2.00	1.00	1.00	.00
M021201	6	14	2.00	1.00	1.00	.00
M044801	7	4	2.00	.00	1.00	1.00
M044901	7	5	2.00	1.00	.00	1.00
M045101	7	7	2.00	1.00	1.00	.00
M045701	7	11	2.00	1.00	1.00	.00
M045801	7	12	2.00	1.00	1.00	.00
M045802	7	12	2.00	1.00	1.00	.00
M045803	7	12	2.00	1.00	1.00	.00
M045804	7	12	2.00	1.00	1.00	.00
M012831	8	7	2.00	1.00	1.00	.00
M013131	8	10	2.00	1.00	1.00	.00
N202831	8	12	2.00	2.00	.00	.00
M052301	9	1	2.00	1.00	1.00	.00
M052401	9	2	2.00	1.00	1.00	.00
M046101	11	2	2.00	.00	2.00	.00
M046501	11	6	2.00	1.00	1.00	.00
M046901	11	10	2.00	1.00	1.00	.00
M047001	11	11	2.00	1.00	1.00	.00
M047101	11	12	2.00	.00	2.00	.00
M047301	11	14	2.00	1.00	1.00	.00
M047801	11	17	2.00	.00	2.00	.00

M053601	12	2	2.00	1.00	1.00	.00
M051301	13	2	2.00	2.00	.00	.00
M051501	13	4	2.00	1.00	1.00	.00
M051701	13	6	2.00	1.00	1.00	.00
M048401	15	4	2.00	1.00	1.00	.00
M048601	15	6	2.00	1.00	1.00	.00
M048801	15	8	2.00	1.00	1.00	.00
M049101	15	10	2.00	1.00	.00	1.00
M032001	16	1	2.00	1.00	.00	1.00
M033201	16	13	2.00	1.00	1.00	.00
M050401	3	6	1.00	1.00	.00	.00
M017601	4	3	1.00	.00	1.00	.00
M019701	6	1	1.00	.00	1.00	.00
M019801	6	2	1.00	.00	1.00	.00
M044501	7	1	1.00	1.00	.00	.00
M013331	8	14	1.00	1.00	.00	.00
M013531	8	16	1.00	1.00	.00	.00
M052801	9	6	1.00	.00	1.00	.00
M046001	11	1	1.00	1.00	.00	.00
M046801	11	9	1.00	.00	1.00	.00
M053501	12	1	1.00	1.00	.00	.00
M053701	12	3	1.00	.00	1.00	.00
M048101	15	1	1.00	1.00	.00	.00
M048501	15	5	1.00	.00	1.00	.00
M048901	15	9	1.00	1.00	.00	.00
M020001	6	4	.00	.00	.00	.00
M048301	15	3	.00	.00	.00	.00

GRADE 12 ITEMS ORDERED BY TOTAL NUMBER OF DESCRIPTORS MAPPED

			Number of Descriptors Consistently Mapped			
NAEPID	BLOCK	ITEM	TOTAL	BASIC	PROFICIENT	ADVANCED
M062401	10	10	8.00	3.00	4.00	1.00
M057101	3	14	7.00	3.00	3.00	1.00
M061907	10	3	6.00	3.00	3.00	.00
M059701	11	12	6.00	4.00	2.00	.00
M059801	11	14	6.00	3.00	3.00	.00
M055701	14	10	6.00	3.00	2.00	1.00
M025401	5	20	5.00	1.00	2.00	2.00
M058101	7	10	5.00	3.00	2.00	.00
M061301	15	6	5.00	3.00	2.00	.00
M056101	3	4	4.00	3.00	1.00	.00
M056901	3	12	4.00	2.00	1.00	1.00
M019001	4	17	4.00	2.00	2.00	.00
M019401	4	21	4.00	2.00	2.00	.00
M021701	6	15	4.00	2.00	2.00	.00
M058001	7	9	4.00	1.00	2.00	1.00
M053301	9	8	4.00	2.00	2.00	.00
M054501	12	8	4.00	3.00	1.00	.00
M056301	3	6	3.00	1.00	2.00	.00
M056501	3	8	3.00	2.00	1.00	.00
M018801	4	15	3.00	3.00	.00	.00

M019301	4	20	3.00	1.00	2.00	.00
M019501	4	22	3.00	.00	2.00	1.00
M024901	5	13	3.00	3.00	.00	.00
M025201	5	16	3.00	2.00	1.00	.00
M025302	5	17	3.00	1.00	1.00	1.00
M021502	6	3	3.00	2.00	1.00	.00
M021101	6	13	3.00	.00	2.00	1.00
M057301	7	2	3.00	1.00	2.00	.00
M057701	7	6	3.00	2.00	1.00	.00
M057801	7	7	3.00	3.00	.00	.00
M057901	7	8	3.00	2.00	1.00	.00
M013031	8	9	3.00	1.00	2.00	.00
M011531	8	15	3.00	2.00	1.00	.00
M012031	8	20	3.00	1.00	2.00	.00
M052701	9	5	3.00	2.00	1.00	.00
M059702	11	13	3.00	.00	3.00	.00
M054001	12	6	3.00	3.00	.00	.00
M054601	12	9	3.00	1.00	2.00	.00
M060701	13	9	3.00	2.00	1.00	.00
M055601	14	9	3.00	3.00	.00	.00
M061101	15	4	3.00	1.00	2.00	.00
M061102	15	4	3.00	1.00	2.00	.00
M061103	15	4	3.00	1.00	2.00	.00
M061104	15	4	3.00	1.00	2.00	.00
M033401	16	15	3.00	2.00	1.00	.00
M033501	16	16	3.00	2.00	1.00	.00
M033801	16	19	3.00	2.00	1.00	.00
M034001	16	21	3.00	3.00	.00	.00
M034101	16	22	3.00	3.00	.00	.00
M055901	3	2	2.00	1.00	1.00	.00
M056001	3	3	2.00	2.00	.00	.00
M056201	3	5	2.00	.00	1.00	1.00
M056401	3	7	2.00	1.00	1.00	.00
M056701	3	11	2.00	2.00	.00	.00
M057001	3	13	2.00	1.00	1.00	.00
M017701	4	4	2.00	1.00	1.00	.00
M018601	4	13	2.00	2.00	.00	.00
M018701	4	14	2.00	2.00	.00	.00
M024001	5	4	2.00	2.00	.00	.00
M024101	5	5	2.00	1.00	1.00	.00
M024201	5	6	2.00	2.00	.00	.00
M024401	5	8	2.00	2.00	.00	.00
M024501	5	9	2.00	.00	2.00	.00
M024701	5	11	2.00	2.00	.00	.00
M021601	6	4	2.00	1.00	1.00	.00
M021602	6	5	2.00	2.00	.00	.00
M020201	6	6	2.00	2.00	.00	.00
M020901	6	11	2.00	.00	2.00	.00
M021001	6	12	2.00	2.00	.00	.00
M021702	6	16	2.00	.00	2.00	.00
M057501	7	4	2.00	2.00	.00	.00
M057601	7	5	2.00	1.00	1.00	.00
M058201	7	11	2.00	2.00	.00	.00
M058301	7	12	2.00	1.00	1.00	.00

M012331	8	2	2.00	2.00	.00	.00
M012631	8	5	2.00	1.00	1.00	.00
M013131	8	10	2.00	1.00	.00	1.00
M013231	8	11	2.00	2.00	.00	.00
M011331	8	13	2.00	2.00	.00	.00
M011831	8	18	2.00	1.00	1.00	.00
M052801	9	6	2.00	1.00	1.00	.00
M053201	9	7	2.00	1.00	1.00	.00
M053401	9	9	2.00	1.00	1.00	.00
M061908	10	4	2.00	1.00	1.00	.00
M061905	10	5	2.00	.00	2.00	.00
M058701	11	3	2.00	.00	2.00	.00
M059401	11	9	2.00	1.00	1.00	.00
M053801	12	4	2.00	1.00	1.00	.00
M060201	13	4	2.00	2.00	.00	.00
M054801	14	2	2.00	.00	2.00	.00
M054901	14	3	2.00	1.00	1.00	.00
M055001	14	4	2.00	2.00	.00	.00
M055101	14	5	2.00	2.00	.00	.00
M055301	14	7	2.00	2.00	.00	.00
M055401	14	8	2.00	2.00	.00	.00
M060901	15	2	2.00	2.00	.00	.00
M061601	15	9	2.00	1.00	1.00	.00
M061801	15	11	2.00	2.00	.00	.00
M033201	16	13	2.00	2.00	.00	.00
M033701	16	18	2.00	2.00	.00	.00
M055801	3	1	1.00	.00	1.00	.00
M017401	4	1	1.00	1.00	.00	.00
M017501	4	2	1.00	1.00	.00	.00
M017801	4	5	1.00	.00	1.00	.00
M017901	4	6	1.00	1.00	.00	.00
M018001	4	7	1.00	.00	1.00	.00
M018101	4	8	1.00	1.00	.00	.00
M018201	4	9	1.00	1.00	.00	.00
M018401	4	11	1.00	1.00	.00	.00
M018501	4	12	1.00	1.00	.00	.00
M019201	4	19	1.00	1.00	.00	.00
M021901	5	1	1.00	1.00	.00	.00
M023901	5	3	1.00	1.00	.00	.00
M024301	5	7	1.00	1.00	.00	.00
M024801	5	12	1.00	1.00	.00	.00
M025001	5	14	1.00	1.00	.00	.00
M025301	5	17	1.00	.00	1.00	.00
M023601	5	19	1.00	.00	1.00	.00
M021401	6	1	1.00	1.00	.00	.00
M020301	6	7	1.00	1.00	.00	.00
M020501	6	9	1.00	1.00	.00	.00
M021201	6	14	1.00	1.00	.00	.00
M021801	6	17	1.00	1.00	.00	.00
M057201	7	1	1.00	1.00	.00	.00
M057401	7	3	1.00	.00	1.00	.00
M057402	7	3	1.00	.00	1.00	.00
M057403	7	3	1.00	.00	1.00	.00
M057404	7	3	1.00	.00	1.00	.00

M012231	8	1	1.00	1.00	.00	.00
M012431	8	3	1.00	1.00	.00	.00
M012531	8	4	1.00	1.00	.00	.00
M012731	8	6	1.00	1.00	.00	.00
M012831	8	7	1.00	1.00	.00	.00
M012931	8	8	1.00	1.00	.00	.00
M011431	8	14	1.00	1.00	.00	.00
M011631	8	16	1.00	.00	1.00	.00
M011731	8	17	1.00	1.00	.00	.00
M011931	8	19	1.00	1.00	.00	.00
M012131	8	21	1.00	1.00	.00	.00
M052501	9	3	1.00	1.00	.00	.00
M052601	9	4	1.00	.00	1.00	.00
M061901	10	1	1.00	.00	1.00	.00
M062001	10	6	1.00	1.00	.00	.00
M062101	10	7	1.00	1.00	.00	.00
M062301	10	9	1.00	.00	.00	1.00
M058901	11	4	1.00	.00	1.00	.00
M059001	11	5	1.00	1.00	.00	.00
M059101	11	6	1.00	1.00	.00	.00
M059201	11	7	1.00	1.00	.00	.00
M059301	11	8	1.00	.00	1.00	.00
M059501	11	10	1.00	1.00	.00	.00
M059601	11	11	1.00	1.00	.00	.00
M053501	12	1	1.00	1.00	.00	.00
M053901	12	5	1.00	1.00	.00	.00
M054401	12	7	1.00	1.00	.00	.00
M059901	13	1	1.00	.00	1.00	.00
M060001	13	2	1.00	1.00	.00	.00
M060101	13	3	1.00	1.00	.00	.00
M060501	13	7	1.00	1.00	.00	.00
M060601	13	8	1.00	1.00	.00	.00
M054701	14	1	1.00	1.00	.00	.00
M055201	14	6	1.00	1.00	.00	.00
M060801	15	1	1.00	1.00	.00	.00
M061201	15	5	1.00	1.00	.00	.00
M061401	15	7	1.00	.00	1.00	.00
M061501	15	8	1.00	.00	.00	1.00
M032001	16	1	1.00	1.00	.00	.00
M032301	16	4	1.00	1.00	.00	.00
M032401	16	5	1.00	1.00	.00	.00
M032501	16	6	1.00	1.00	.00	.00
M032601	16	7	1.00	1.00	.00	.00
M032701	16	8	1.00	1.00	.00	.00
M033001	16	11	1.00	1.00	.00	.00
M033901	16	20	1.00	1.00	.00	.00
M056601	3	9	.00	.00	.00	.00
M056801	3	10	.00	.00	.00	.00
M017601	4	3	.00	.00	.00	.00
M018301	4	10	.00	.00	.00	.00
M018901	4	16	.00	.00	.00	.00
M019101	4	18	.00	.00	.00	.00
M022001	5	2	.00	.00	.00	.00
M024601	5	10	.00	.00	.00	.00

M025101	5	15	.00	.00	.00	.00
M023501	5	18	.00	.00	.00	.00
M021501	6	2	.00	.00	.00	.00
M020401	6	8	.00	.00	.00	.00
M020801	6	10	.00	.00	.00	.00
M058401	7	13	.00	.00	.00	.00
N202831	8	12	.00	.00	.00	.00
M052301	9	1	.00	.00	.00	.00
M052401	9	2	.00	.00	.00	.00
M061904	10	2	.00	.00	.00	.00
M062201	10	8	.00	.00	.00	.00
M058501	11	1	.00	.00	.00	.00
M058601	11	2	.00	.00	.00	.00
M053601	12	2	.00	.00	.00	.00
M053701	12	3	.00	.00	.00	.00
M060301	13	5	.00	.00	.00	.00
M060401	13	6	.00	.00	.00	.00
M061001	15	3	.00	.00	.00	.00
M061701	15	10	.00	.00	.00	.00
M032101	16	2	.00	.00	.00	.00
M032201	16	3	.00	.00	.00	.00
M032801	16	9	.00	.00	.00	.00
M032901	16	10	.00	.00	.00	.00
M033101	16	12	.00	.00	.00	.00
M033301	16	14	.00	.00	.00	.00
M033601	16	17	.00	.00	.00	.00

Appendix I

Assignment of Items to Multiple and Single Levels by Item Format and Content

Table I.1

Number of Items Classified to Single and Multiple Achievement Levels by Item Format, Grade 4

Level	Multiple-choice	Constructed response	Extended constructed response	Total
Not Classified	19	9	0	28
Basic	6	0	0	6
Proficient	73	36	0	109
Basic & Proficient	20	2	0	22
Advanced	0	1	0	1
Basic & Advanced	0	0	0	0
Proficient & Advanced	1	6	5	12
Basic, Proficient, & Advanced	0	0	0	0
Total	119	54	5	178

Table I.2

Number of Items Classified to Single and Multiple Achievement Levels by Item Format, Grade 8

Level	Multiple-choice	Constructed response	Extended constructed response	Total
Not Classified	1	1	0	2
Basic	9	4	0	13
Proficient	8	3	0	11
Basic & Proficient	68	39	3	110
Advanced	0	0	0	0
Basic & Advanced	9	1	0	10
Proficient & Advanced	6	1	0	7
Basic, Proficient, & Advanced	45	10	5	58
Total	146	59	6	211

Table I.3

Number of Items Classified to Single and Multiple Achievement Levels by Item Format, Grade 12

Level	Multiple-choice	Constructed response	Extended constructed response	Total
Not Classified	25	9	0	34
Basic	63	24	1	88
Proficient	17	7	0	24
Basic & Proficient	35	11	3	49
Advanced	2	0	0	2
Basic & Advanced	0	1	0	1
Proficient & Advanced	2	1	0	3
Basic, Proficient, & Advanced	1	4	2	7
Total	145	57	6	208

Table I.4

Number of Items Classified to Single and Multiple Achievement Levels by Item Content, Grade 4

Level	Numbers& Operations	Measure-ment	Geometry	Data Analysis	Algebra& Functions	Estima-tion	Total
Not classified	12	2	11	2	1	0	28
Basic	1	0	0	0	0	5	6
Proficient	41	24	13	13	12	6	109
Basic & Proficient	7	3	0	0	3	9	22
Advanced	0	0	0	1	0	0	1
Basic & Advanced	0	0	0	0	0	0	0
Proficient & Advanced	2	2	3	4	1	0	12
Basic, Proficient, & Advanced	0	0	0	0	0	0	0
Total	63	31	27	20	17	20	178

Table I.5

Number of Items Classified to Single and Multiple Achievement Levels by Item Content, Grade 8

Level	Numbers& Operations	Measure-ment	Geometry	Data Analysis	Algebra& Functions	Estima-tion	Total
Not classified	2	0	0	0	0	0	2
Basic	6	3	0	2	2	5	13
Proficient	2	2	3	3	1	0	11
Basic & Proficient	21	17	20	23	20	9	110
Advanced	0	0	0	0	0	0	0
Basic & Advanced	4	0	0	0	3	3	10
Proficient & Advanced	2	4	1	0	0	0	7
Basic, Proficient, & Advanced	21	6	12	6	3	10	58
Total	58	32	36	34	29	22	211

Table I.6

Number of Items Classified to Single and Multiple Achievement Levels by Item Content, Grade 12

Level	Numbers& Operations	Measure-ment	Geometry	Data Analysis	Algebra& Functions	Estima-tion	Total
Not classified	11	4	2	6	4	7	34
Basic	24	18	15	1	18	12	88
Proficient	2	1	10	9	2	0	24
Basic & Proficient	5	6	9	11	15	3	49
Advanced	0	0	0	0	2	0	2
Basic & Advanced	0	0	0	1	0	0	1
Proficient & Advanced	0	0	0	1	2	0	3
Basic, Proficient, & Advanced	2	0	1	0	4	0	7
Total	44	29	37	29	47	22	208

Appendix J

Mean, Median, Minimum and Maximum P-Values for Students Scoring at Each Achievement Level on Sets of Items Mapped to Each Descriptor

GRADE 4 (178 Items)

P-VALUES FOR BELOW BASIC STUDENTS, GRADE 4

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	105	.304	.263	.030	.862
D2	None	34	.262	.208	.008	.744
D6	None	101	.259	.233	.001	.750
D8	None	24	.131	.069	.001	.442
D1A	Basic	10	.320	.289	.142	.830
D1B	Basic	14	.346	.351	.142	.595
D2A	Basic	4	.328	.318	.165	.513
D3	Basic	177	.306	.262	.001	.862
D6A	Basic	1	.750	.750	.750	.750
D1C	Prof	56	.305	.264	.031	.862
D1D	Prof	12	.325	.287	.105	.617
D1E	Prof	10	.311	.332	.055	.663
D2B	Prof	13	.234	.184	.008	.649
D4	Prof	87	.248	.208	.001	.853
D6B	Prof	46	.254	.235	.037	.649
D7	Prof	70	.235	.212	.001	.750
D8A	Prof	6	.029	.031	.001	.067
D8B	Prof	8	.030	.031	.001	.070
D5	Advanced	1	.023	.023	.023	.023
D6C	Advanced	1	.376	.376	.376	.376
D8C	Advanced	3	.139	.070	.004	.343
D8D	Advanced	11	.124	.067	.001	.435

P-VALUES FOR BASIC STUDENTS, GRADE 4

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	105	.514	.493	.158	.937
D2	None	34	.408	.373	.051	.921
D6	None	101	.457	.452	.019	.949
D8	None	24	.343	.260	.019	.765
D1A	Basic	10	.502	.472	.162	.911
D1B	Basic	14	.485	.468	.187	.841
D2A	Basic	4	.465	.408	.300	.742
D3	Basic	177	.505	.491	.019	.980
D6A	Basic	1	.948	.948	.948	.948
D1C	Prof	56	.526	.526	.158	.937
D1D	Prof	12	.481	.456	.209	.841

D1E	Prof	10	.493	.496	.221	.855
D2B	Prof	13	.360	.275	.087	.727
D4	Prof	87	.430	.418	.019	.980
D6B	Prof	46	.435	.449	.051	.762
D7	Prof	70	.432	.410	.019	.949
D8A	Prof	6	.141	.153	.019	.228
D8B	Prof	8	.156	.143	.019	.334
D5	Advanced	1	.246	.246	.246	.246
D6C	Advanced	1	.698	.698	.698	.698
D8C	Advanced	3	.344	.334	.083	.613
D8D	Advanced	11	.311	.228	.019	.724

P-VALUES FOR PROFICIENT STUDENTS, GRADE 4

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	105	.722	.748	.255	.982
D2	None	34	.616	.573	.166	.984
D6	None	101	.671	.695	.163	.984
D8	None	24	.576	.535	.163	.943
D1A	Basic	10	.697	.738	.362	.907
D1B	Basic	14	.670	.667	.315	.961
D2A	Basic	4	.655	.566	.538	.951
D3	Basic	177	.703	.748	.163	.997
D6A	Basic	1	.965	.965	.965	.965
D1C	Prof	56	.735	.792	.255	.982
D1D	Prof	12	.680	.706	.315	.961
D1E	Prof	10	.696	.707	.369	.908
D2B	Prof	13	.569	.533	.252	.912
D4	Prof	87	.639	.660	.163	.994
D6B	Prof	46	.640	.682	.166	.906
D7	Prof	70	.649	.675	.163	.984
D8A	Prof	6	.373	.384	.163	.537
D8B	Prof	8	.393	.390	.163	.602
D5	Advanced	1	.614	.614	.614	.614
D6C	Advanced	1	.890	.890	.890	.890
D8C	Advanced	3	.556	.602	.288	.779
D8D	Advanced	11	.525	.534	.163	.898

P-VALUES FOR ADVANCED STUDENTS, GRADE 4

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	105	.878	.924	.427	1.000
D2	None	34	.845	.866	.539	1.000
D6	None	101	.858	.920	.427	1.000
D8	None	24	.774	.848	.236	1.000
D1A	Basic	10	.884	.933	.661	.987
D1B	Basic	14	.865	.898	.585	1.000

D2A	Basic	4	.868	.865	.742	1.000
D3	Basic	177	.866	.918	.236	1.000
D6A	Basic	1	.997	.997	.997	.997
D1C	Prof	56	.886	.942	.427	1.000
D1D	Prof	12	.872	.922	.654	1.000
D1E	Prof	10	.865	.900	.654	.971
D2B	Prof	13	.840	.898	.539	.973
D4	Prof	87	.836	.898	.416	1.000
D6B	Prof	46	.843	.922	.427	1.000
D7	Prof	70	.839	.890	.427	1.000
D8A	Prof	6	.672	.672	.416	.898
D8B	Prof	8	.672	.675	.416	.898
D5	Advanced	1	.916	.916	.916	.916
D6C	Advanced	1	.959	.959	.959	.959
D8C	Advanced	3	.730	.680	.593	.918
D8D	Advanced	11	.711	.680	.416	.945

GRADE 8 (211 Items)

P-VALUES FOR BELOW BASIC STUDENTS, GRADE 8

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D6	None	63	.303	.252	.001	.861
D13	None	17	.441	.453	.016	.772
D14	None	17	.305	.216	.004	.754
D22	None	18	.131	.044	.000	.572
D1	Basic	78	.396	.343	.0001	.906
D4	Basic	15	.246	.195	.003	.636
D6A	Basic	48	.303	.248	.001	.792
D8	Basic	81	.323	.271	.001	.914
D9	Basic	89	.291	.256	.0001	.914
D10	Basic	42	.285	.248	.0001	.862
D12	Basic	2	.208	.208	.156	.260
D15	Basic	194	.353	.303	.0001	.914
D2	Prof	108	.319	.265	.0001	.853
D3	Prof	10	.247	.205	.086	.506
D5	Prof	3	.096	.111	.003	.175
D6B	Prof	49	.292	.231	.001	.861
D7	Prof	77	.365	.358	.001	.914
D13A	Prof	9	.424	.381	.156	.752
D13B	Prof	3	.390	.381	.016	.772
D14A	Prof	13	.307	.179	.004	.754
D14B	Prof	6	.617	.713	.381	.754
D14C	Prof	2	.326	.326	.271	.381
D16	Prof	6	.087	.067	.000	.213
D18	Prof	1	.572	.572	.572	.572
D19	Prof	6	.119	.038	.000	.386
D22A	Prof	2	.098	.098	.024	.172
D22B	Prof	8	.156	.105	.000	.381
D22C	Prof	6	.110	.014	.000	.379

D22E	Prof	9	.012	.004	.000	.051
D6C	Advanced	16	.230	.213	.005	.572
D11	Advanced	2	.412	.412	.371	.453
D17	Advanced	1	.010	.010	.010	.010
D20	Advanced	1	.735	.735	.735	.735
D21	Advanced	50	.448	.459	.075	.906
D22D	Advanced	15	.147	.038	.0001	.572

P-VALUES FOR BASIC STUDENTS, GRADE 8

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D6	None	63	.506	.501	.018	.965
D13	None	17	.658	.731	.103	.954
D14	None	17	.511	.551	.070	.977
D22	None	18	.290	.223	.018	.798
D1	Basic	78	.582	.633	.029	.966
D4	Basic	15	.457	.373	.022	.890
D6A	Basic	48	.501	.512	.027	.928
D8	Basic	81	.530	.541	.018	.987
D9	Basic	89	.492	.523	.018	.987
D10	Basic	42	.465	.406	.022	.957
D12	Basic	2	.323	.323	.188	.457
D15	Basic	194	.568	.599	.018	.987
D2	Prof	108	.531	.542	.018	.966
D3	Prof	10	.427	.335	.152	.752
D5	Prof	3	.178	.229	.022	.283
D6B	Prof	49	.492	.482	.027	.965
D7	Prof	77	.580	.623	.018	.987
D13A	Prof	9	.623	.654	.180	.902
D13B	Prof	3	.570	.654	.103	.954
D14A	Prof	13	.493	.329	.070	.977
D14B	Prof	6	.897	.956	.654	.977
D14C	Prof	2	.666	.666	.654	.679
D16	Prof	6	.198	.111	.022	.628
D18	Prof	1	.798	.798	.798	.798
D19	Prof	6	.241	.162	.027	.542
D22A	Prof	2	.331	.331	.111	.551
D22B	Prof	8	.359	.387	.029	.686
D22C	Prof	6	.244	.090	.027	.686
D22E	Prof	9	.080	.037	.018	.232
D6C	Advanced	16	.362	.287	.037	.839
D11	Advanced	2	.720	.720	.673	.766
D17	Advanced	1	.071	.071	.071	.071
D20	Advanced	1	.928	.928	.928	.928
D21	Advanced	50	.661	.717	.159	.962
D22D	Advanced	15	.309	.232	.018	.798

P-VALUES FOR PROFICIENT STUDENTS, GRADE 8

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D6	None	63	.676	.727	.085	.996
D13	None	17	.803	.917	.196	.990
D14	None	17	.675	.735	.196	.998
D22	None	18	.457	.501	.085	.895
D1	Basic	78	.750	.829	.100	.996
D4	Basic	15	.691	.744	.134	.995
D6A	Basic	48	.669	.710	.085	.989
D8	Basic	81	.702	.764	.127	.996
D9	Basic	89	.684	.763	.100	.998
D10	Basic	42	.637	.685	.100	.989
D12	Basic	2	.603	.603	.363	.843
D15	Basic	194	.742	.829	.085	.998
D2	Prof	108	.726	.811	.100	.993
D3	Prof	10	.668	.685	.368	.934
D5	Prof	3	.398	.490	.134	.569
D6B	Prof	49	.666	.693	.085	.996
D7	Prof	77	.727	.845	.085	.996
D13A	Prof	9	.750	.804	.196	.980
D13B	Prof	3	.699	.735	.374	.990
D14A	Prof	13	.673	.721	.269	.998
D14B	Prof	6	.950	.989	.735	.998
D14C	Prof	2	.821	.821	.735	.907
D16	Prof	6	.371	.354	.100	.867
D18	Prof	1	.895	.895	.895	.895
D19	Prof	6	.417	.390	.100	.726
D22A	Prof	2	.528	.528	.259	.796
D22B	Prof	8	.532	.604	.100	.839
D22C	Prof	6	.409	.290	.100	.839
D22E	Prof	9	.237	.160	.085	.521
D6C	Advanced	16	.539	.502	.085	.930
D11	Advanced	2	.919	.919	.890	.939
D17	Advanced	1	.160	.160	.160	.160
D20	Advanced	1	.973	.973	.973	.973
D21	Advanced	50	.812	.896	.341	.996
D22D	Advanced	15	.467	.481	.085	.895

P-VALUES FOR ADVANCED STUDENTS, GRADE 8

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D6	None	63	.833	.904	.349	1.000
D13	None	17	.889	.981	.154	1.000
D14	None	17	.829	.943	.154	1.000
D22	None	18	.658	.730	.240	.947
D1	Basic	78	.869	.940	.240	1.000
D4	Basic	15	.868	.956	.442	1.000
D6A	Basic	48	.825	.906	.349	1.000

D8	Basic	81	.850	.930	.384	1.000
D9	Basic	89	.843	.930	.240	1.000
D10	Basic	42	.797	.831	.240	1.000
D12	Basic	2	.812	.812	.649	.976
D15	Basic	194	.871	.945	.154	1.000
D2	Prof	108	.866	.946	.240	1.000
D3	Prof	10	.832	.868	.551	1.000
D5	Prof	3	.694	.742	.442	.899
D6B	Prof	49	.830	.904	.349	1.000
D7	Prof	77	.855	.944	.349	1.000
D13A	Prof	9	.832	.954	.154	1.000
D13B	Prof	3	.843	.795	.738	.996
D14A	Prof	13	.862	.943	.557	1.000
D14B	Prof	6	.964	.998	.795	1.000
D14C	Prof	2	.880	.880	.795	.964
D16	Prof	6	.639	.710	.240	.937
D18	Prof	1	.944	.944	.944	.944
D19	Prof	6	.628	.671	.240	.849
D22A	Prof	2	.783	.783	.618	.947
D22B	Prof	8	.726	.796	.240	.947
D22C	Prof	6	.635	.707	.240	.822
D22E	Prof	9	.515	.442	.240	.796
D6C	Advanced	16	.757	.740	.349	.984
D11	Advanced	2	.975	.975	.972	.979
D17	Advanced	1	.421	.421	.421	.421
D20	Advanced	1	1.000	1.000	1.000	1.000
D21	Advanced	50	.901	.962	.458	1.000
D22D	Advanced	15	.664	.764	.240	.947

GRADE 12 (208 Items)

P-VALUE FOR BELOW BASIC STUDENTS, GRADE 12

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	70	.306	.281	.000	.827
D2	None	51	.251	.199	.000	.902
D3	None	9	.174	.091	.000	.458
D4	None	29	.378	.375	.004	.821
D14	None	38	.199	.048	.000	.821
D1A	Basic	22	.238	.264	.0003	.620
D1B	Basic	46	.272	.228	.000	.820
D1F	Basic	1	.296	.296	.296	.296
D2A	Basic	36	.205	.193	.000	.510
D2D	Basic	5	.266	.167	.002	.720
D4A	Basic	1	.821	.821	.821	.821
D4B	Basic	9	.319	.329	.004	.821
D4E	Basic	2	.435	.435	.252	.618
D8	Basic	5	.364	.332	.248	.506
D6	Basic	19	.404	.470	.000	.857
D9	Basic	7	.455	.506	.248	.606
D11	Basic	53	.401	.340	.000	.904

D14	Basic	8	.021	.001	.000	.116
D14	Basic	8	.044	.023	.000	.116
D1C	Prof	12	.358	.311	.000	.820
D1D	Prof	13	.332	.325	.0004	.827
D1E	Prof	10	.437	.437	.044	.820
D2B	Prof	17	.161	.153	.000	.458
D2C	Prof	9	.201	.204	.029	.458
D3A	Prof	3	.159	.091	.068	.319
D3C	Prof	1	.091	.091	.091	.091
D4C	Prof	7	.265	.252	.052	.499
D4D	Prof	18	.385	.484	.004	.658
D10	Prof	7	.528	.506	.248	.817
D14D	Prof	16	.170	.048	.000	.690
D14E	Prof	1	.094	.094	.094	.094
D14F	Prof	8	.111	.023	.000	.490
D3B	Advanced	2	.275	.275	.091	.458
D3D	Advanced	1	.089	.089	.089	.089
D3E	Advanced	4	.211	.193	.000	.458
D5	Advanced	1	.091	.091	.091	.091
D13	Advanced	1	.114	.114	.114	.114
D14C	Advanced	5	.110	.044	.000	.408

P-VALUES FOR BASIC STUDENTS, GRADE 12

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	70	.511	.537	.001	.959
D2	None	51	.463	.429	.002	.978
D3	None	9	.352	.456	.051	.667
D4	None	29	.586	.688	.034	.955
D14	None	38	.395	.293	.000	.955
D1A	Basic	22	.403	.361	.025	.858
D1B	Basic	46	.464	.354	.001	.959
D1F	Basic	1	.567	.567	.567	.567
D2A	Basic	36	.428	.356	.010	.919
D2D	Basic	5	.409	.335	.002	.974
D4A	Basic	1	.955	.955	.955	.955
D4B	Basic	9	.484	.359	.034	.955
D4E	Basic	2	.605	.605	.281	.929
D8	Basic	5	.574	.679	.291	.741
D6	Basic	19	.616	.679	.000	.964
D9	Basic	7	.583	.676	.291	.804
D11	Basic	53	.609	.575	.033	.986
D14A	Basic	8	.080	.027	.000	.351
D14B	Basic	8	.138	.093	.002	.351
D1C	Prof	12	.510	.530	.008	.918
D1D	Prof	13	.578	.661	.033	.930
D1E	Prof	10	.680	.804	.142	.918
D2B	Prof	17	.327	.289	.010	.682
D2C	Prof	9	.491	.498	.287	.682
D3A	Prof	3	.352	.456	.137	.463
D3C	Prof	1	.137	.137	.137	.137

D4C	Prof	7	.403	.359	.194	.809
D4D	Prof	18	.599	.726	.034	.948
D10	Prof	7	.708	.681	.291	.958
D14D	Prof	16	.315	.170	.002	.887
D14E	Prof	1	.285	.285	.285	.285
D14F	Prof	8	.227	.076	.002	.710
D3B	Advanced	2	.368	.368	.141	.595
D3D	Advanced	1	.133	.133	.133	.133
D3E	Advanced	4	.391	.459	.051	.595
D5	Advanced	1	.137	.137	.137	.137
D13	Advanced	1	.334	.334	.334	.334
D14C	Advanced	5	.244	.142	.002	.809

P-VALUE FOR PROFICIENT STUDENTS, GRADE 12

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	70	.732	.804	.162	.983
D2	None	51	.733	.806	.068	.998
D3	None	9	.711	.777	.478	.941
D4	None	29	.735	.854	.271	.993
D14	None	38	.622	.658	.045	.992
D1A	Basic	22	.641	.701	.201	.961
D1B	Basic	46	.718	.774	.162	.982
D1F	Basic	1	.750	.750	.750	.750
D2A	Basic	36	.734	.818	.139	.998
D2D	Basic	5	.577	.695	.068	.987
D4A	Basic	1	.983	.983	.983	.983
D4B	Basic	9	.659	.647	.305	.983
D4E	Basic	2	.621	.621	.271	.971
D8	Basic	5	.811	.867	.622	.925
D6	Basic	19	.763	.854	.045	.992
D9	Basic	7	.756	.794	.491	.954
D11	Basic	53	.810	.868	.265	.993
D14A	Basic	8	.266	.188	.045	.677
D14B	Basic	8	.328	.215	.068	.677
D1C	Prof	12	.670	.778	.230	.912
D1D	Prof	13	.779	.864	.201	.981
D1E	Prof	10	.802	.898	.201	.975
D2B	Prof	17	.663	.714	.139	.968
D2C	Prof	9	.808	.860	.508	.968
D3A	Prof	3	.717	.777	.516	.860
D3C	Prof	1	.516	.516	.516	.516
D4C	Prof	7	.573	.619	.271	.956
D4D	Prof	18	.746	.850	.305	.993
D10	Prof	7	.866	.925	.622	.983
D14D	Prof	16	.496	.504	.068	.963
D14E	Prof	1	.560	.560	.560	.560
D14F	Prof	8	.381	.215	.068	.888
D3B	Advanced	2	.728	.728	.552	.904
D3D	Advanced	1	.567	.567	.567	.567
D3E	Advanced	4	.755	.818	.478	.904

D5	Advanced	1	.516	.516	.516	.516
D13	Advanced	1	.547	.547	.547	.547
D14C	Advanced	5	.388	.201	.068	.956

P-VALUES FOR ADVANCED STUDENTS, GRADE 12

DESCRIPTOR	LEVEL	N	MEAN	MEDIAN	MIN	MAX
D1	None	70	.878	.940	.308	1.000
D2	None	51	.885	.945	.308	1.000
D3	None	9	.901	.924	.762	.991
D4	None	29	.849	.945	.331	1.000
D14	None	38	.801	.904	.308	1.000
D1A	Basic	22	.838	.892	.372	.997
D1B	Basic	46	.889	.940	.514	1.000
D1F	Basic	1	.784	.784	.784	.784
D2A	Basic	36	.893	.947	.418	1.000
D2D	Basic	5	.769	.882	.457	1.000
D4A	Basic	1	.988	.988	.988	.988
D4B	Basic	9	.819	.916	.562	.988
D4E	Basic	2	.654	.654	.331	.977
D8	Basic	5	.953	.947	.931	.996
D6	Basic	19	.866	.949	.319	1.000
D9	Basic	7	.904	.923	.801	.977
D11	Basic	53	.921	.972	.418	1.000
D14A	Basic	8	.552	.550	.308	.946
D14B	Basic	8	.597	.551	.308	.893
D1C	Prof	12	.824	.859	.464	.987
D1D	Prof	13	.870	.949	.372	.968
D1E	Prof	10	.864	.961	.372	1.000
D2B	Prof	17	.866	.929	.549	1.000
D2C	Prof	9	.950	.971	.755	1.000
D3A	Prof	3	.853	.855	.762	.944
D3C	Prof	1	.762	.762	.762	.762
D4C	Prof	7	.740	.867	.331	.970
D4D	Prof	18	.869	.936	.562	1.000
D10	Prof	7	.953	.982	.828	.996
D14D	Prof	16	.697	.690	.308	.996
D14E	Prof	1	.850	.850	.850	.850
D14F	Prof	8	.623	.577	.372	.987
D3B	Advanced	2	.903	.903	.815	.991
D3D	Advanced	1	.924	.924	.924	.924
D3E	Advanced	4	.930	.936	.855	.991
D5	Advanced	1	.762	.762	.762	.762
D13	Advanced	1	.618	.618	.618	.618
D14C	Advanced	5	.654	.578	.372	.970

Appendix K

Median *P*-Values for Students in Each Level on Sets of Items Mapped to Single and Multiple Levels

Table K.1

Median *P*-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Single and Multiple Levels, Grade 4

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	28	.318	.635	.792	.935
Basic	6	.424	.471	.673	.851
Proficient	109	.240	.479	.743	.920
Basic & Proficient	22	.302	.472	.705	.941
Advanced	1	.435	.724	.898	.945
Basic & Advanced	0				
Proficient & Advanced	12	.049	.237	.535	.763
Basic, Proficient, & Advanced	0				

Table K.2

Mean *P*-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Single and Multiple Levels, Grade 8

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	2	.601	.774	.895	.953
Basic	13	.410	.843	.919	.947
Proficient	11	.453	.808	.876	.961
Basic & Proficient	110	.274	.553	.803	.947
Advanced	0				
Basic & Advanced	10	.495	.700	.897	.949
Proficient & Advanced	7	.561	.792	.915	.985
Basic, Proficient, & Advanced	58	.263	.547	.807	.910

Table K.3

Median *P*-Values for Below Basic, Basic, Proficient and Advanced Students on Subsets of Items Assigned to Single and Multiple Levels, Grade 12

Highest level of descriptor to which item was mapped	# of items	Level of students			
		Below basic	Basic	Proficient	Advanced
Not classified	34	.315	.613	.848	.959
Basic	88	.354	.667	.879	.968
Proficient	24	.470	.699	.862	.952
Basic & Proficient	49	.248	.364	.750	.905
Advanced	2	.090	.137	.559	.869
Basic & Advanced	1	.114	.334	.547	.618
Proficient & Advanced	3	.319	.456	.777	.854
Basic, Proficient, & Advanced	7	.044	.142	.478	.893

Appendix L

Single Level Classifications of Items Administered to Multiple Grades

	Grade 12		
Grade 8	Basic	Proficient	Advanced
Basic	16	3	0
Proficient	3	2	0
Advanced	0	0	0

	Grade 8		
Grade 8	Basic	Proficient	Advanced
Basic	0	1	2
Proficient	2	7	16
Advanced	0	0	4

	Grade 12		
Grade 4	Basic	Proficient	Advanced
Basic	3	0	0
Proficient	26	2	0
Advanced	6	0	0

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☑ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").